

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Baker Examiner #: 76550 Date: 10/28  
 An Unit: 1645 Phone Number 30 8-8886 Serial Number: 091868, 987  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Search seq 14

against nucleic acid ds  
except pending

Ph Ar Sy D3  
Socher cluck 13  
Road T of Lon 20  
Vanilla Collection

1 20  
3h 3h  
4  
5p  
6p  
6h

## STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>D. Schneider</u>	NA Sequence (#) <u>1</u>	STN: _____
Searcher Phone #: <u>308-4292</u>	AA Sequence (#) _____	Dialog: _____
Searcher Location: <u>CM 6403</u>	Structure (#) _____	Questel/Orbit: _____
Date Searcher Picked Up: <u>10/29/03</u>	Bibliographic _____	Dr. Linn: _____
Date Completed: <u>10/29/03</u>	Litigation _____	Lexis Nexis: _____
Searcher Prep & Review Time: <u>12</u>	Fulltext _____	Sequence Systems: <u>CompuGen</u>
Client Prep Time: _____	Patent Family _____	WWW Internet: _____
Online Time: <u>6</u>	Other _____	Other (specify): _____

BEST AVAILABLE COPY

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 94 Seconds  
(without alignments)  
2591.952 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKPNHASVSK.....SLYIAPPLLEFMRKNSK 552

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spo01/US09868987/runat 28102003\_152602\_23852/app query.fasta\_1.711  
-DB=Issued\_Patents\_NA -QFWI=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987@cgn 1.1.56 @runat 28102003\_152602\_23852 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT\_DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/6C COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/6D COMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2730	98.3	1230025	4	US-09-198-452A-1
C 2	719	25.9	1830121	4	US-09-557-884-1
C 3	719	25.9	1830121	4	US-09-643-990A-1
C 4	678.5	24.4	2211	3	US-09-462-844-1
C 5	672	24.2	6321	4	US-09-221-017B-311
C 6	646.5	23.3	2301	4	US-09-134-001C-852
C 7	512	18.4	4403765	3	US-09-103-840A-2
C 8	509	18.3	4411529	3	US-09-103-840A-1
C 9	483.5	17.4	40429	4	US-08-311-731A-125
C 10	425	15.3	1398	4	US-09-252-991A-6413
C 11	418	15.0	1947	4	US-09-328-352-1027
C 12	417	15.0	1890	4	US-09-252-991A-6330

FILE COPY

FILE COPY

ALIGNMENTS

RESULT 1

Sequence 1, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
TITLE OF INVENTION: and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198.452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 1  
LENGTH: 1230025  
TYPE: DNA  
ORGANISM: Chlamydia pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(15000)  
NAME/KEY: misc feature  
LOCATION: (15001)..(30000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (30001)..(45000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (45001)..(60000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (60001)..(75000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (75001)..(90000)  
OTHER INFORMATION: n=a or c or g or t

Sequence 6245, Ap  
Sequence 6054, Ap  
Sequence 6217, Ap  
Sequence 6138, Ap  
Sequence 1243, Ap  
Sequence 1, Appli  
Sequence 4040, Ap  
Sequence 4006, Ap  
Sequence 3906, Ap  
Sequence 1348, Ap  
Sequence 897, App  
Sequence 2168, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1459, A  
Sequence 14791, A  
Sequence 829, App  
Sequence 191, App  
Sequence 14790, A  
Sequence 14540, A  
Sequence 14823, A  
Sequence 443, App  
Sequence 1012, Ap  
Sequence 418, App  
Sequence 10659, A  
Sequence 10414, A  
Sequence 10504, A  
Sequence 14948, A  
Sequence 14688, A  
Sequence 15066, A  
Sequence 319, App  
Sequence 102, App



```

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

```

## Alignment Scores:

```

Pred. No.: 0 Length: 1230025
Score: 2730.00 Matches: 551
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 98.27% Indels: 3
DB: 4 Gaps: 0

```

US-09-868-987-14 (1-552) x US-09-198-452A-1 (1-1230025)

```

Qy 1 MetValSerSerProfileLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db 661906 ATGGTCAGCAGGCGCTATTTTAAACGTCCCATTTGAAATAATCATGCCAGTGTCTCAGGGAAA 661847
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 661846 TTTACCACCGTGAAGTAGGACAACTCGCCTCAGATTTTAAATCTTGGAGCGATGCTTTT 661787
Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
Db 661786 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTTGGGAAAAAACAATGTACA 661727
Qy 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
Db 661726 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTAAGCGGTATAT 661667
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
Db 661666 TATAGATTTGGAGGCGTCATCGCTTCGGAGCTGTTCTTCTCAATCTTTTGCTTATCTGG 661607
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 661606 GCAGCTCTACAGTATTGGATGGCGACCTCACCTTTGTACAGGACTCGCTGGGATGTTCTT 661547
Qy 121 AlaMetGlyMetAlaValAspAlaValLeuValPheGluArgIleArgGluPhe 140
Db 661546 GCTATGGGATGGCCGTAGATGCAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTT 661487
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 661486 TTATGCTCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCC 661427
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 661426 ATTTTGTATCTAACTTGAACAGTATTTGGGCTCAGCATTTTGGCTCAGCACTTCTTTCTTCCATAGATACA 661367
Qy 181 GlyProIleLysGlyPheAlaLeuThrIleLeuGlyIlePheSerSerMetPheThr 200
Db 661366 GGGCCCTATTAAAGGGTTTGGCTTGGACATTTTGGGAAATTTTCTCTTCAATGTTTACG 661307
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220

```

```

Db 661306 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG 661247
Qy 221 LeuHisMetMetAenLysPheValGlyIleIleYehiSapPheLeuArgGlyCysLysLys 240
Db 661246 TTGCATATGATGAATAAGTTCGTGGGATGAAGCATGATTTCTTGAGAGGATGCAAAAAA 661187
Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 661186 CTTTGGGCTGTTCTCGAAGTGTTTTCTTTAGTTGGCTGTCTCGGTTGGAGCC 661127
Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280
Db 661126 TGGAAATCCGTTTGGGAATGGATTTTAAAGAGGGTATGCTTTTACCTTTAAATCCAAA 661067
Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db 661066 GAGCATGGCATCAGCATGTTGCTCAATGCTGGCAAGTTTGTCATTAACACTACAGNA 661007
Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLys 320
Db 661006 GCTGCTCTTCTCTAGAGACTTCCGTATTCAACATTTGSGATCTTCAGAAAAAGATCAAA 660947
Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThr-LysGlnIleArgAlaSerLeuLeuLys 340
Db 660946 ATCTATTTTGTAGTATAAGCTTTAAGCTATATAAGCAGATACAGCAGCTCTCTCTCAA 660887
Qy 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValAtqAsnArgProArgPheL 360
Db 660886 ATTAACGATCATGAGCTGGCGTTAGCTGTGGATTTGTTCAGAAACAGGCTAGATTTTC 660827
Qy 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379
Db 660826 TCTACGGAAACTCTAAACGAAACCGCAAAATTTTGGTCAAAGGTAAAGCAGCAACTATCG 660767
Qy 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399
Db 660766 AAGAAATGCGTTATCAGGCGACCATCGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 660707
Qy 400 TyrValSerLeuArgPheGluTyrAlaPheSerAlaValCysAlaLeuIleHis 419
Db 660706 TATGTAGTTTGGCGCTTGAATGGCAATATGCTTTCAGTGCGGTATGCGCTTTTAATTCAT 660647
Qy 420 AspLeuLeuAlaThrCysAlaValPheIleAlaHisPhePheLeuLysLysIleGln 439
Db 660646 GACCTTTTGGCTACTCTGCGAGTCTTGTATATACACATCTTCTTTTGAAGAAATTCAA 660587
Qy 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThr 459
Db 660586 ATAGATTTGCAAGCCATTTGGTGTCTTAATGACTGTATTGGGGTATTCATTAAACAATACT 660527
Qy 460 LeuIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db 660526 TTGATCATTTTGTATGATTCGTGAAGATCGCAAGCAACCTGTTTACCCCTTATGCTAT 660467
Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db 660466 GTTTTAGTTAATGATGCCCTTCAAAGACGTTTACGCCGACGGTAAATGACAAACAGCTACA 660407
Qy 500 ThrLeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAla 519
Db 660406 ACTCTACAGTTTGTAAATGCTTTGTTTATAGCGGCTCCTCTGTCTTTTAAATTTTGA 660347
Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db 660346 TTTATATGACCATAGGATTTCTTAGGAATCTTATGCTCTCTTTATATATGACCACT 660287
Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db 660286 CTGTTGTTGTTATGTCCTGCTTAAAGAAATCGCTCAAAA 660248

```

RESULT 2

US-09-557-884-1/c

; Sequence 1, Application US/09557884

Patent No. 6506581  
 GENERAL INFORMATION:  
 APPLICANT: Fleischmann et al.  
 TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
 NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3 1/2 inch diskette  
 COMPUTER: Dell Pentium  
 OPERATING SYSTEM: MS DOS v6.22  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/557,884  
 FILING DATE: 25-Apr-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/476,102  
 FILING DATE: JUN-5-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michelle S. Marks  
 REGISTRATION NUMBER: 41,971  
 REFERENCE/DOCKET NUMBER: PB186P3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-309-8504  
 TELEFAX: 301-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1

Alignment Scores:  
 Pred. No.: 1,37e-74 Length: 1830121  
 Score: 719.00 Matches: 186  
 Percent Similarity: 55.69% Conservative: 127  
 Best Local Similarity: 33.10% Mismatches: 174  
 Query Match: 25.88% Indels: 75  
 DB: 4 Gaps: 18

US-09-868-987-14 (1-552) x US-09-557-884-1 (1-1830121)

Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22  
 Db 268552 AATGTTGCAACAATTCAGGACGTTTGGTTCTTAATTTCCAAATTTACTGGTGTGTGATAGC 268493  
 Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42  
 Db 268492 ATTGCGGAGACAAATAATCTTTTACCTTATTGAAATCTGGTCATTAAATGACCAATT 268433  
 Qy 43 GluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGlnGly 62  
 Db 268432 CAAATTTGCAAGAACCGCAAAATTTGGCCCATCATTTAGTTCGCCAAACGTTAGACGAGG 268373  
 Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyTyArg 82  
 Db 268372 ATTAATGCGAGTCTTTGGGAGATTAGTTGCTGTTATTGCTTTTATGTTTACTACAAA 268313  
 Qy 83 PheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeuLeuIleTrpAlaAla 102  
 Db 268312 ATGTTTGGTGTGATGCAAGTTTGCACTGTTATTAAATATCGTATTATTTGGGATTA 268253  
 Qy 103 LeuGlnTyLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121

Db 268252 ATGCTATTATTTACCCGCGGACACATTTCAATGCCGGTATTCCGGGTATCGTTTAACT 268193  
 Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 268192 TTAGGTATGTGCTAGATGCAATGATTTGATTTTGAACGTATTAAAGAGAAATTCGT 268133  
 Qy 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyThrLysAlaPheGlyAlaIle 161  
 Db 268132 AATGGTCGTTCATTCAGCAAGCCATTAAATGAAGGTATTAAACGGCGCATTTACTTCTATT 268073  
 Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181  
 Db 268072 TTGTATGCAAACTTAACCAATCTTAACCGCAATTAITCTATACCGGTAGGAACAGC 268013  
 Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201  
 Db 268012 CCAATTCAGGTTTTCGATTAACGCTTCACITTTGGTGTTCGATTTCTATGTTTACCGCG 267953  
 Qy 202 LeuPheMetThr-----LysPhe 207  
 Db 267952 ATTACAGGAACTCGCGCATTTAGTTAATGCCCTTTACGCTGGTAAACAACTTAAAAAATTA 267893  
 Qy 208 Phe-PheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMetMetAs 225  
 Db 267892 TTAATTTAGCGCGGAAATGATGAAACTTTTACAAAGACATAAAGACGGACATTTTATCCG 267833  
 Qy 225 nLysPheValGlyIleLys-----HisAspPheLeuArgGlyCysLysLys 240  
 Db 267832 TGAATTCATGATGGATAAAGCTCCGCTTCCATTGCAATTTATGAAAGTCGTAATTT 267773  
 Qy 240 sleuTrpAlaValSerGlySerValPheLeuLeuGly-----CysValAlaLeuG1 257  
 Db 267772 GGGTTATATATATCCGCACATTTTTCATGATGTAATTTCTCTATTTTATTTATACCAAGG 267713  
 Qy 257 yPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPheThrPh 277  
 Db 267712 ATTTAAC---TGG-----GGCTTAGATTTTACTGGTGGAGTGGTATTTGATAC 267668  
 Qy 277 eAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValIleHisLys 297  
 Db 267667 TCACTTCTCGCAGTCC-----GCTAACCTTGAACAAATTCGTAGT-----AA 267626  
 Qy 297 sleuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerG1 317  
 Db 267625 ACTTCACGAAATGGAATTTGAAAGCCCA-----ATTGTACAAACCCACAGGATCGTTTCA 267572  
 Qy 317 uLysIleLysIleTyTyPheSerAspLysAlaLeuSerTyThrLysGlnIleArgAlaSe 337  
 Db 267571 CGATGTGATGATT-----CGTTTACCTGCAAG 267545  
 Qy 337 rleuLeuLysLeuThrIleMetSerTrpArgTyCysGlyIleValValArgAsnArgPr 357  
 Db 267544 TAAATAATGATTTCTACCATTT-----GGTGAACACGCTCAAAAGTATGTACAGATGTAGA 267491  
 Qy 357 cArgPheLeuTyGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerLys 377  
 Db 267490 TAAAGACATT-----CAAAATTCGAGTATTGAGTTC-----GTTGCCCAAA 267449  
 Qy 377 sleuSerLysLysMetArgTyGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle1 397  
 Db 267448 TCTTGGTCAAGAAATTAGCACAGGTGGGTATATCGGACTTACGACATTTAGCAATGGT 267389  
 Qy 397 eleuLeuTyTrpValSerLeuArgPheGluTrpGlnTyAlaPheSerAlaValCysAlaLe 417  
 Db 267388 GCTTATTTATGGGGTCCACGTTTGAATGGCGTTTAGCGTTTGGCAGTATCGTCTCT 267329  
 Qy 417 uIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLys 437  
 Db 267328 TCGCACACGCTCAATTATTACGTTAGGGGTA-----TTCTCTGCATTACA 267284  
 Qy 437 sileGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyTrSerLeuAs 457

Db 267283 AATTGAAATGATCTTACTTTTTCGCGAGCGATTTTATCTGTGGTGTACTCCATCAA 267224  
 Qy 457 nasnThrLeuIlePheAspAqlleArgGluasp--ArgGlnAlaAsnLeuPheTh 476  
 Db 267223 CGATAGTATGTGGTATTTGACCGGGTTCGTGAAATTTCCGAAAAATTAGACGATTGGA 267164  
 Qy 476 rProMethisValLeuAlaAsnAspAlaLeuGlnLysThrPheSerArgThrValMetTh 496  
 Db 267163 TAGCATTCATATT--ATTGATATTTCTTAACGCAAACTTTATCAAGAACTATCATTTAC 267107  
 Qy 496 rThraLThrLeuSerValLeuMetLeuMetLeuPheleGlyGlySerSerValPh 516  
 Db 267106 TTCGGTTACTACATAGTTGTGCGTGCATGTCCTTCTTCTTGGTGGTCTTCCATTCA 267047  
 Qy 516 eAsnPheAlaPheIleMetThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrl 536  
 Db 267046 TAACTTTTCACTTGCTTACTCGTAGGTATGGATTGGTACTTATCTCGATTTTGT 266987  
 Qy 536 eAla 537  
 Db 266986 TGCC 266983

## RESULT 3

US-09-643-990A-1/c

; Sequence 1, Application US/09643990A

; Patent No. 6528289

; GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Mark D. Adams

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Haemophilus influenzae Rd Genome, Fragments  
 ; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESS: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville,

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

; OPERATING SYSTEM: MS DOS v6.22

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/643,990A

; FILING DATE: 23-AUG-2000

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/487,429

; FILING DATE: 1995-06-07

; APPLICATION NUMBER: 08/426,787

; FILING DATE: 1995-04-21

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PB186P1C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-610-5790

; TELEFAX: 310-309-8439

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1830121 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:  
 Pred. No.: 1.37e-74 Length: 1830121  
 Score: 719.00 Matches: 186  
 Percent Similarity: 55.69% Conservative: 127  
 Best Local Similarity: 33.10% Mismatches: 174  
 Query Match: 25.88% Indels: 75  
 DB: 4 Gaps: 18

US-09-868-987-14 (1-552) x US-09-643-990A-1 (1-1830121)

Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22  
 Db 268552 AATGTTCAACAATTTCAAGACGTTTGGTTCTAAATTTCCAAATTTACTGGTGTGATAGC 268493

Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42  
 Db 268492 ATTGCGGAGACACATAATCTTTCTACCTTATGAAATCTGGTGCATTAATTTGCACCAATT 268433

Qy 43 GluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGlnGly 62  
 Db 268432 CAAATTCGTAAGAACGACCAATTTGGCCCATTCATTAGTGGCAAAACGTAGACGAAGG 268373

Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyArg 82  
 Db 268372 ATTAATCGGAGTCTTTGGGATTAGTTGCTGTATTATGCCCTTTATGTTGTTTACTACAA 268313

Qy 83 PheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuIleTrpAlaAla 102  
 Db 268312 ATGTTTGGTGTGATTGCAAGTTTGCACCTTGTATTATTAATATCTGTTACTTGTGGGATTA 268253

Qy 103 LeuGlnTyLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121  
 Db 268252 ATGTCCTATTTTACCCGGCGGACACTTTCAATCGCGGTATTGGCGGTATCGTTTAACT 268193

Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 268192 TTAGGTATGTAGTAGATGCGAATGTATTGATTTTGAACGTATTAAAGAAGAAATTCGT 268133

Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyThrLysAlaPheGlyAlaIle 161  
 Db 268132 AATGTCGTTCAATTCAGCAAGCCATTAAAGAGTTATAACGGCGGATTTACTTCTATT 268073

Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThrGly 181  
 Db 268072 TTTGATGCAAACTTAACCAATCTTAACCGCAATTTATCTTATACGGGTAGGAACAGGC 268013

Qy 182 ProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThrAla 201  
 Db 268012 CCAATTCGAAGTTTGGGATTACGCTTTCACTGTGTGTCGATTTCTATGTTTACCGCG 267953

Qy 202 LeuPheMetThr-----LysPhe 207  
 Db 267952 ATTACAGGAATCGCGCATTTAGTTAATGCCCTTTACGGTGTAAACAACCTTAAATAATTA 267893

Qy 208 Phe-PheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMetMetAs 225  
 Db 267892 TTAATTTAGCGGGAATGATGAAACTTTTACAAAAGATAAAGACGACATTTATCCG 267833

Qy 225 nLysPheValGlyIleLys-----HisAspPheLeuArgGlyLysLysLys 240  
 Db 267832 TGAATCAATGGGATAAAGCTCCCGTTCCCAATTCACCTGAATTTATGAAGTGGCGTAAAT 267773

Qy 240 sLeuTrpAlaValSerGlySerValPheLeuLeuGly-----CysValAlaLeuGly 257  
 Db 267772 GGGTTATATATTCGCGCACTTTTGATGGTAAATTTCTCTATTTTATTTATACCAAGG 267713

Qy 257 yPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPheThrPh 277  
 Db 267712 ATTTAAC---TGG-----GGCTTAGATTTTACTGCGAGTGGTATTGTATAC 267668

Qy 277 eAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLys 297  
 Db 267667 TCACTTTCGCGAGTCC-----GCTAACCTTTGAACAAATTCGTAGT-----AA 267626







D	b		1875	GGATCTTTCGCGTTCTCGCTCGCTTGGTTATPCCGATGTGTACATGTGTCTGGCTTAC	18181
Q	y		82	ArgPhe--GlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp	100
D	b		1815	GTTTCTTGCCGGGTCTTATCGCAACCGCGCATGTATTAAACAGCTTCTCACATTG	1756
Q	y		101	AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu	120
D	b		1755	GGCGTATTGGCTTCTTCATCGCGCTGTGCACCCTCTCGGTATCGCAGGTTTGGTGTG	1696
Q	y		121	AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleAArgGluGluPhe	140
D	b		1695	ACGCTGGGTATGGCTGTGGATGCCAACGACTATTATCTCGAGCGTATCAAAGAAGAGCTT	1636
Q	y		141	LeuLeuSerGinSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla	160
D	b		1635	CGTGCCGTAAGACTCCGATTCGTGCGCTTAGCGATGGTTATGCCAACGCTTCTCTGCC	1576
Q	y		161	IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr	180
D	b		1575	ATCTTCGACTCGAAGCTTACGACTATTATTACC GGATATCTCTTATCTCTACGGGACG	1516
Q	y		181	GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr	200
D	b		1515	GGCGCGATTCCGGTTTTTGCACATACGTTGATTATTCGGTCTTATCGCTTCTTTCATTACG	1456
Q	y		201	AlaLeuPheMetThrLysPhePhePhe-MetLeuTrpMetAsnLysThrGlnHisThrGl	220
D	b		1455	GCTGCTCTTTCGACTCGTATCGTCTTTCGAGAACTGGCGAAAAGGTCGTTGGATAAG	1396
Q	y		220	nLeuHisMetMetAsnLysPheValGlyIleLys-----HisAspPheLeuAr	236
D	b		1395	ATTACATTCACTACGACGATTACTCGCAATCTCTCTGTCATCCCTCATACAAC-ATCTT	1337
Q	y		236	gGlyCysLysLysLeu---TrpAlaValSerGlySerValPheLeuLeuGlyCysValAl	255
D	b		1336	GGGTAAGCGCAAGACCGGCTTTATCATCTCCGGTGATTATCATGCTTTTGGGACTTATAGC	1277
Q	y		255	a-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGl	272
D	b		1276	TTCATTACATTCGGTCTCAATAG-----GGTATTGAATTCCTCCGGAGG	1232
Q	y		272	Y-----TyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMe	290
D	b		1231	ACGTAACCTACGTAGTTAAATTC-----GACCAGCGTGATCTCCGAAGCC---GT	1184
Q	y		290	tArgGlyLysValValHisLysLeuGlnGlu-----AlaGlyLeuSe	304
D	b		1183	TCGTTCCGCGCTGTCTTCTCCCCTGCGAGAAAAAGGATTGGTTTACCTCCGTACTION	1124
Q	y		304	rSerArgAspPheArgIleGlnThr-----PheGlySerSerGluLysIleLy	320
D	b		1123	AGGACAGAGTGGCTTATATCTCGAACTATAGATCCAGGAGGAAAGCGAAGAACTGA	1064
Q	y		320	sileTyrPheSerAspLysAlaLeuSer-----TyrThrLysGlnIleAr	335
D	b		1063	AGCAGAGATTACTGACAAAATGGTATCAGACGCTGAAAGGTTTCTACACCCAGCAGCTTAC	1004
Q	y		335	gAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAs	355
D	b		1003	TGCTGATCAGTCTCTTGACAATATCATAGCTCTCAG-----	967
Q	y		355	nArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSe	375
D	b		966	-----AAAGTAAG	959
Q	y		375	rSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAl	395
D	b		958	TCCAGGATGTGCGATGACATCAGGAGTGCTATTTGGGCTGTGCTGTATCATCAT	899
Q	y		395	alleIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAla----	413
D	b		898	CTTCATGGCCATTATATCTGATTCGCTTCGCT---GACATTCTTCTCTCGCGGGT	842

QY 414 -----ValCysAlaLeuHisAspLeuLeuAlaThrCysAlaValLeuPhe-- 429  
Db 841 ATTGCTATCTGTGGCGCTACTACATTCTGCATTATTCTCTGTATGCGTGTCTGTGGAA 782  
QY 430 -IleAlaHisPheLeuLysLysIleGlnIleaspLeuGlnAlaIleGlyAlaLeuMe 449  
Db 781 GATTCTGCCCTTC-----ACCATGGAGATCGATCAGAACTTCATCGCTGCTATTCT 731  
QY 449 tThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAs 469  
Db 730 GGCTATCATCGGTACTCGCTCAATGACACCGTGGTGTATTGTACCGGTATCCGAGAGAC 671  
QY 469 pArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysTh 489  
Db 670 GATGAAATGTGACCCCAACAGAGATCGTATCAGGTGATCAACGATGCCCTTAATTCAAC 611  
QY 489 rPheSerArgThrValMetThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPhe 509  
Db 610 ATTGGTGGCAATTAATACGTCCTTGACTACGTTTATCGTTATCGTTATGTTGGTAATCTTCAT 551  
QY 509 eIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuG1 529  
Db 550 CTTTGGAGGTCTACGATGCGTAGTTTCACGTTCTCGATCCTGCTCGGTATCGTTATCGG 491  
QY 529 yThrLeuSerSerLeuTyrlleAlaProProLeuLeuLeuPheMetValArgLysGluAs 549  
Db 490 TACATACTCTACGCTCTTTGTTGCTACACCC--CTTSCCTACGAGATCCAAAAGCGCAA 434  
QY 549 nArgSerLys 552  
Db 433 GCTCAACAAA 424  
RESULT 6  
US-09-134-001C-852  
; Sequence 852, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GPC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 852  
; LENGTH: 2301  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-852

Alignment Scores:	
Pred. No.:	1.62e-70
Score:	646.50
Length:	2301
Matches:	164
Conservative:	129
Percent Similarity:	51.22%
Best Local Similarity:	28.67%
Mismatches:	200
Query Match:	23.37%
Indels:	79
DB:	4
Gaps:	12

US-09-868-987-14 (1-552) x US-09-134-001C-852 (1-2301)

Qy	2	ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe	21
Db	625	GTAGACCAACCTATTAACTTAGTAGGTGTGAATTTCAGTGCCTTCATGGGAAAAA	684
Qy	22	ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal	41
Db	685	GGTGTTGAAGAGCGCAACAAATAGCTGAGTTTAAATCGCGGCTCAATTCACGATTGAT	744



Qy	42	ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThrGln	61
Db	745	TTAAAGAAATTTACTCTAACTCTGTTGGTGCAAAATTCGTCAAGATGCTCTTGATAAG	804
Qy	62	GlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTyr	81
Db	805	ACCATGTTTCATCAATTGTAGGTATAGCATTAATTTATTTATTTATGCTTGTTGTTCTAT	864
Qy	82	ArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuIleTyrAla	101
Db	865	CGITTCCTGGTTAGTTGCAATCATTCCTTAACCACTTATATTTATTTAACTTTAGTC	924
Qy	102	AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla	121
Db	925	GCATTCAAATTCATATCAGGTGATTAACTCTACCTGGATTGGCGCATTAGTTAGGT	984
Qy	122	MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu	141
Db	985	GTAGGTAGGCTGTCGATGCAATATCATATGTATGACGTATTAAGAATGAACCTAAGA	1044
Qy	142	LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle	161
Db	1045	ATTGGACGCACGCTTAAACAAGCGTATTCAAAGCAATAAAGATTCATTTAACTATA	1104
Qy	162	PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly	181
Db	1105	TTTGATTCCAACTTAAACAATGTCATCGCTGCAGCTGCTGCTTTCTTTTGGAGAAAGT	1164
Qy	182	ProIleLysPheAlaLeuThrLeuIleLysGlyIlePheSerSerMetPheThrAla	201
Db	1165	TCAGTCAAAAGCGTTCGCAACCATGTTACTCTTAGGTATTTTAATGATATTTGTAACGCA	1224
Qy	202	LeuPheMetThrLysPhePheMetLeu	211
Db	1225	GTATTCCTTCAAGAGGGTGTATCATTTACTGCTATCTTCAAACTCTTTAAAAACAA	1284
Qy	212	---TrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle	230
Db	1285	TACTGGTTATTTGGTGTAAAGAGATGGATAGACATCATATTAATGAAGGTAAAGATGTA	1344
Qy	231	LysHisAsp	241
Db	1345	---CATGATTTAAAAACATCATATGAAGGTTAAACTTTGTATAATTAGCTAAGCCACTT	1401
Qy	242	TrpAlaValSerGlySerValPheLeuGlyCysValAlaLeuGlyPheGlyAlaTyr	261
Db	1402	ATTTCACTTAGTATTTTAATGTAAATTTATGGTTTGATTCATTTCCAAATTTAAATTA	1461
Qy	262	AsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGlu	281
Db	1462	AAC---TTAGTATTGATTTCTCATCCGAAACAAGACAGACATATTCATACTAAAAAT	1515
Qy	282	HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla	301
Db	1516	-----GCTATAACAACAAGCACAGGTTCAGAAAACTGTAAAAATCATAGTT	1557
Qy	302	GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle	321
Db	1558	GGATTGAACACAGAT-----CAATACAGATTAATGTTAGTGGAAATAAA	1602
Qy	322	TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeu	341
Db	1603	-----AATGCCACAGTTCAGTTTAAAAAAGATTTCACAGT---	1638
Qy	342	ThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyr	361
Db	1639	-----GAG	1641
Qy	362	GlyAsnSerLysArgAsnAlaLysPheTrpSerLys	373
Db	1642	GAAGCAACATAATTAAGTGTGAAGGTGAATCTGAATTTGGAGATATATCCACAAATTAAT	1701
Qy	374	---ValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly	392

```

Db      1702  ACGGTTTCACCTCTCATAGCCCAAGACTAGCTAAAAATCCTGTAACTGCATTAACTT 1761
Qy      393  AlaleuAlailelleuLeuTyValSerLeuArgPheGluTrpGlnTyAlapheSer 412
Db      1762  GCTTCTATAGGCATTATATCTATGTTCTACTAAGATTTCGAATGGCGTGATGGGTCTATCT 1821
Qy      413  AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheilleAlaHis 432
Db      1822  TCTGTACTTCGATTATTTACATGACGTATTT-----ATCATCATTTGCAATC 1866
Qy      433  PhePheLeuLysylsIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452
Db      1867  TTTAGTTTGTGTAGATTAGAAGTAGATTTAACAATTATTCAGCAGATTAATACTATCGTT 1926
Qy      453  GlyTySerLeuAsnAenThrLeuIleIlePheAspArgIleArzGluAsp---ArgGln 471
Db      1927  GGTTATTCAATCAATGATACATCAATCGTAACCTTCGACCGTGTTCGAGAAAATTCGCATAA 1986
Qy      472  AlaAsnLeuPheThr-----ProMetHisValLeuValAenAspAlaLeuGlnLys 488
Db      1987  GTTAAAGTAATTACGCATCTAGTCAAAATTCATGATGATAGTCAACCGCTCTATTAGCAA 2046
Qy      489  ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508
Db      2047  ACTATGACACGTTCTATTAAATACAGTGTTCAGCTGTAGTTGTAGTTGCCAATATTA 2106
Qy      509  PheileGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528
Db      2107  ATATTAGTGCACCAACAAATATTAAATTTCTTTTACCATTACTAATTTGATTATATCT 2166
Qy      529  GlyThrLeuSerSerLeuTyIleAlaProLeu 540
Db      2167  GGTTATTCTCGICAAATTTTCATTGCTGTACCATT 2202

RESULT 7
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURES:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:          6,23e+48           Length:          4403765
Score:             512.00              Matches:           161
Percent Similarity: 42.58%              Conservative:     123
Best Local Similarity: 24.14%            Mismatches:       221
Query Match:       18.43%               Indels:           162
DB:                  3                   Gaps:             18

US-09-868-987-14 (1-552) x US-09-103-840A-2 (1-4403765)

Qy      2 ValSerProfileuAenValProLeuLysAenHisAlaSerValSerGlyLys---- 20
      ::: ||| :::::::::::::::::::: |::| |::| |::|

```

```
Db 2910893 CTGGACTCGCAGGTTGTCAAGTCCAGGATCCAGGAGCGATCCCGCGCGCGAACC 2910834
Qy 21 -----PheThrHisArgGluValSerLysLeuAlaSerAsp 32
Db 2910833 CAGATCAGCGGTGGGATCCCGCAATTCACCGCGGACCGCACGCGAGCTCGCAACGTC 2910774
Qy 33 LeuLysSerGlyAlaMetSerPheValProGluValLeuSerGluGluThrIleSerSer 52
Db 2910773 CTCAAGTACGGGTGCTCGCTCCGCTGCTCCGAAACCGTGGGAGCCAAACCGTTTCGGCG 2910714
Qy 53 AspleuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72
Db 2910713 ACCTTGGGTTGTATCGCTCGCGCGGCGCATGATCGCGGTGCGATCGCGCTCTGTTG 2910654
Qy 73 LeuIleValLeuMetSerValTyrThrArgPheGlyGly-----Val 86
Db 2910653 GTGTGGTGTATTCGCTCTACTACCGAGTGTGGGACTGCTCAGCGCCTTGTGCTA 2910594
Qy 87 IleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaAlaLeuGlnTyrLeu 106
Db 2910593 GTGGCTTCGCTCAATGGTTTTTTCGATCTGCTGGTCTTTGGGC-----CGATATATC 2910540
Qy 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaVal 126
Db 2910539 AACTACACCCCTAGATCTGGCTGGTATTTCGGGCGCTGATCATCGGTATCGGACCAACCGCC 2910480
Qy 127 AspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeu 146
Db 2910479 GACTCGTTCGTTGGTGTTCGAAACGCATCAACAGCAGATCCCGAAGCGGTTCGTTTC 2910420
Qy 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeu 166
Db 2910419 CGGTCCGAGTCCCGCGGTTGGCGGTCCCGAGAGCATCGTGTGGGCAACGCC 2910360
Qy 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186
Db 2910359 GTCACTTTCCTGGCGCGTGGCTGTACTTTCTGGCGATCGTCAAGTGAAGGGTTC 2910300
Qy 187 AlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLys 206
Db 2910299 GCGTTCACCCCTG-----GGCTCACACGATCCTGGACCTCGTCGTG 2910258
Qy 207 PhePhePheMetLeuTrp-----MetAsnLys 215
Db 2910257 GTGTCTTGTGACGTGGCGGTGGTGTATCTGCGTCCAAATCGTCTGCTGGCCAAAG 2910198
Qy 216 ThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIleLys-----231
Db 2910197 CCGGCATACAACGGCTGGGAGCGGTGCAGCAGGTGCACGCGAAGCCCGCGGCATGGCT 2910138
Qy 231 -----231
Db 2910137 CGGACGGGACGGGATAGCCGATGGCGTCAAAAGCAAGACCGCGCGACGAGGCC 2910078
Qy 231 -----231
Db 2910077 ACGTGCGGGTGCNACTCACCGAGGCCACCGAGNGTCTGTGGCGGACCGAGCGGCAC 2910018
Qy 232 -----HisAspPheLeuArg-----236
Db 2910017 AGCACACAGACACCGCTCGAAGCTGGGGCACCATAGTTCCTTTCGCGCTGTACACC 2909958
Qy 237 -----GlyCysLysLysLeuTrp---AlaValSerGlySer 247
Db 2909957 GGCACCGGGCGTTGTAGGTAGTGGGACCGCGCGGTGTGGTGTGGCGTGAAGCGGTGCG 2909898
Qy 248 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet 267
Db 2909897 ATCGTCGCGTGGCCATCGCCAGCATGCTCTTTCGTTGGGTTC-----ACCTTCGCGCATC 2909844
Qy 268 AspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVal 287
Db 2909843 GATTTCAAAGCGGTACACCGGTGTCTGTTTC-----CCGCGTGGCAGACCCCAAGTC 2909793
```

```
Qy 288 AlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGlyLeuSerSerArgAsp 307
Db 2909792 GCGCAGGTGAAGACGCTCTACTACCGGGCCCTC-----GCCAGTGAAGCCCGCATCG 2909742
Qy 308 PheArgGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAla 327
Db 2909741 GTGGTATCTCGGGCGGTGCTTCGGGACCGTGCAGATTCGTCGGAGACGCTGACC 2909682
Qy 328 LeuSerTyrThrLysGlnIleArgAlaSerLeuLysLeuThrIleMetSerTrpArg 347
Db 2909681 AGCGACGAGCGCCAAAGCTGGCGGACGCCCTATTTCGAGGCCTTC-----2909637
Qy 348 TyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsn 367
Db 2909636 -----GGGCCCCAAGGAACCGAGCTCAGCCCGACAGCAG 2909601
Qy 368 AlaLysPheTrpSerLysValSerLysLysMetArgTyrGlnAlaThr 387
Db 2909600 GCATCAGGACTCGCGGTGTTCGGAGACCTGGGGCGGTTCAGATCACCAAGAGCGGTG 2909541
Qy 388 IleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTrp 407
Db 2909540 ATCGCGCTGGTGTGTTCTGTGTGTGTGCGCTCTACATTACCGTGCCTACGAGCGC 2909481
Qy 408 GlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaVal 427
Db 2909480 TACATGACCATCTCGCGGATAACGGCCATGCTCTTCGACCTGACCGTCCCGCGCGTG 2909421
Qy 428 LeuPheIleAlaHisPheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAla 447
Db 2909420 TATTCCTCGTGGTTC-----GAAGTCACCCCGCCCGCTCATCGGC 2909376
Qy 448 LeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArg 467
Db 2909375 CTGCTGACCATCTCGGGTTCCTCGCTCTATGACACCGTCTATGTTCCACAAGGTGCGAG 2909316
Qy 468 GluAsp-----ArgGlnAlaAsnLeu 474
Db 2909315 GAGAACACCATCGGTCTCCAGCACACACCCCGCGCACCTTCGCCGAGCAGGCCAACCTG 2909256
Qy 475 PheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVal 494
Db 2909255 -----GCCATCAACCAACGCTTCATCGTGCATC 2909226
Qy 495 MetThrThrAlaThr---ThrLeuSerValLeuLeuMetLeuLeu-----Phe 509
Db 2909225 AACACACCGCTCATCGCGGTGTTCGGGTGTTCGGCGTGTATGCTGTGTGGCGGTCTGGCTG 2909166
Qy 510 IleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGly 529
Db 2909165 CTGGGCGTCCGACCTCTCAAGACCTGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2909106
Qy 530 ThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPhe-----543
Db 2909105 ACCTACTGCTCAATCTTCTTCGCCACCGCGTGTGGT-TACCTCGTGAGCGCAGGAG 2909046
Qy 544 MetValArgLysGluAsnArg 550
Db 2909045 TTGGTGGCGCAACACACCGCT 2909025
```

RESULT 8  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: PRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00





Db 12513 AGGCGGCACACCGGTGTCCATG---CCTGTCTCCCGGAGTGGGGGGACAGCGCA 12457  
Qy 290 etArgGlyValValHisLysLeuGlnGluAlaGlyLeuSerSerArgPheArgI 310  
Db 12456 TTGAGTTGCCAGGTTCGCCATGCTTCAAGAACACCTTGGCAGATATCCGGAATCGG 12397  
Qy 310 leGlnThrPheGlySerSerGluLysIleLeuLysIleTyPheSerAspLysAlaLeuSer- 329  
Db 12396 TGGTGGTTGGTAACGGCGGTGAGCAACGGTGGGATCAGTTCGAAACGGTGTCCA 12337  
Qy 330 -----TyThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgT 348  
Db 12336 ATGACACAGCTGCAAGCTGGCTTACGCGCTTGTGACGCATTC----- 12293  
Qy 348 yrCysGlyIleValValArgAenArgProArgPheLeuTyGlyAsnSerLysArgAenA 368  
Db 12292 -----GGACCCAAAGGCGCGGAGCTTAAGCTTAGAAGCAGG 12256  
Qy 368 laLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyGlnAlaThrI 388  
Db 12255 CAATCAGCAGCGCGGTATCGGAGACCTGGGGCGGTGAGTACCAAGAGGTGTGA 12196  
Qy 388 leGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyValSerLeuArgPheGluTrpG 408  
Db 12195 TCGCTCTTGTGTGTCTGT 12136  
Qy 408 InTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValL 428  
Db 12135 ATATGGCTATCTCCGACGTACTACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12076  
Qy 428 euPheIleAlaHisPheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAlaL 448  
Db 12075 ACTCGTGTGGCTTC-----GAAGTCACTCCGCGCACTGTCTCGCT 12031  
Qy 448 euMetThrValLeuGlyTySerLeuAsnThrLeuIlePheAspArgIleArgG 468  
Db 12030 TGTTCACGATCTCGGTCTCTCTCTACGACACCGTTATCGTGTGTGTGTGTGTGTGT 11971  
Qy 468 luAsp-----ArgGlnAlaLeuP 475  
Db 11970 AGAATACCCACGGTTTCAGCAGCACACTCGCGGCACCTTCGCTGAGCAGGCTAATCTG- 11912  
Qy 475 heThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValM 495  
Db 11911 -----GCGATCAATCAGACCTTCATGCGTTCGATCA 11881  
Qy 495 etThrAlaThrThrLeuSerValLeuLeuMetLeuLeu-----PheI 510  
Db 11880 ACACGAGCCTGATCTCGGTACTACCGGTGTGTGGCGTGTGTGGTGTGTGTGTGTGT 11821  
Qy 510 leGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuGlyT 530  
Db 11820 TGGGCGTGGTACGCTCAAGGACCTGGCATTAAGTCAGCTGTGTGGCATCATTTGCGCA 11761  
Qy 530 hrLeuSerSerLeuTyIleAlaProProLeuLeuLeu 542  
Db 11760 CTAATCGTGTGATCTCTTTGCAACTCCACTGTGTGTGTC 11723

## RESULT 10

US-09-252-991A-6413  
; Sequence 6413, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,789  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6413  
; LENGTH: 1398  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6413  
Alignment Scores:  
Pred. No.: 4,46e-43 Length: 1398  
Score: 425.00 Matches: 96  
Percent Similarity: 59.09% Conservative: 47  
Best Local Similarity: 39.67% Mismatches: 79  
Query Match: 15.30% Indels: 20  
DB: 4 Gaps: 3  
US-09-868-987-14 (1-552) x US-09-252-991A-6413 (1-1398)  
Qy 1 MetValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 727 ATCATCAGCTGGCGACCATCCAGTCCGCTGGCAACACAGTTCCGTATATCACCGGCTC 786  
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 787 GATGGCCCCGGCGAGTCTCCGAGCTGGCCCTCTGCTGCTGCGCGCGCTGCTGCTGCG 846  
Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 847 CCATGTACTTCCCGGAGAACCGACCATCCGCGCGAGCTGGGGCGGCAACATCGCC 906  
Qy 61 GlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTy 80  
Db 907 AAGGTATCATGATCGCTCTGGGCGATCTCTCGTATCGCTGTGTTCATCATCATCATC 966  
Qy 81 TyArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuLeuLeuLeuLeuLeu 100  
Db 967 TACGCTTCTTCGCGGTGATCGCGACCTGGCTGGCTTCAACATGGTGTGCTGCTGCTG 1026  
Qy 101 AlaAlaLeuGlnTyLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1027 GCGCTGATGTGATCTCTCGCGGCGACCTGACCTCCGCGGCGATCGCGGTATCGTCTG 1086  
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140  
Db 1087 ACCATGGGATGGCGGTGGAGCGCAACGTGCTGATCTCTCGCGGATACGCGAGAACTG 1146  
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyThrLysAlaPheGlyAla 160  
Db 1147 GCCAATGGCATGTCCGTCACAAAGGCGATCCAGAGGCTTCAACCGTGTGCTTCCCGCG 1206  
Qy 161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Db 1207 ATCTTCAGCGCAACCTGACCTCGCTGCTGGTGGCGGCATCTATACGCCATGGGACC 1266  
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThr 200  
Db 1267 GGCCCGGTGAAGATTGGCTGACCATGCTGCGTGGCATCATCATCATCATCATCATCAT 1326  
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
Db 1327 GCCATATGTGTACCCGCG-----GGCGGTGACTTC-----AAGAAG 1386  
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysValLys 240  
Db 1345 ---GCAATGTGTCAACCTGATCTTCGCGC---GGCGGTGACTTC-----AAGAAG 1386  
Qy 241 LeuTrp 242  
Db 1387 CTGTGG 1392  
RESULT 11  
US-09-328-352-1027  
; Sequence 1027, Application US/09328352  
; Patent No. 6562958

## ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GT99-03FA  
; CURRENT APPLICATION NUMBER: US/09/328.352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1027  
; LENGTH: 1947  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1027

## Alignment Scores:

Pred. No.: 6e-42 Length: 1947  
Score: 418.00 Matches: 85  
Percent Similarity: 64.68% Conservative: 56  
Best Local Similarity: 38.99% Mismatches: 77  
Query Match: 15.05% Indels: 0  
DB: 4 Gaps: 0

US-09-868-987-14 (1-552) x US-09-328-352-1027 (1-1947)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 1273 GTGATTAATGTCGCAACTGTTCAAGCCGTTTAGGTTCAACTTTCGGTATTACTGGTTTA 1332  
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 1333 AGTTACCTCAAGAAGCATCTGAACCTTCGCTTAATGCTGCTGCTGCTGCTGCTGCTGCG 1392  
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 1393 CCAATGTACTTCGTGAAGACGCTGTTGGGTCCAAAGCCTTGGTCAAGAAACATTGAT 1452  
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 1453 AAAGGTGTACTATCAACGCAAACTCGGTTCTTGTGTTGCAATCTGGATGGTCTGATTC 1512  
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeuLeuIleTyr 100  
Db 1513 TTCCGTCATTGCTGTTAAATCGGCACTTTGCACTTGTATTAACTTGGCAATGATTTTA 1572  
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1573 ACCATTATGTCGTGATAGTGCTTCCTCACCTTACCGGGTATTGCGGGTATCGTCATT 1632  
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValPheGluArgIleArgGluGluPhe 140  
Db 1633 ACCATTGTTATGCGCGTCGACGCAACGTACTGATGTGAACGTATTCGAGAAGAAATG 1692  
Qy 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
Db 1693 CTCTGGGGGGCCCTCACCTAAACAGGCCAATGTGGCGGGTATTGATCGAGCCTATAACACC 1752  
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Db 1753 ATTTTCGACTCGAACTTAAACAGCTTCCTTGTGGTCATTCTGTTTGAATCGGTACA 1812  
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Db 1813 GGCCCGATTAAAGGTTTCGCGGTGACATTAATGATCGGTATTATTGCTCAATGTTTACT 1872  
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHis 218  
Db 1873 GCCATTACAGTACGCGTGGATTTGACAAATCATTTATGGTAAACGCGGTAAC 1926

## RESULT 12

US-09-252-991A-6330  
; Sequence 6330, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6330  
; LENGTH: 1890  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6330

## Alignment Scores:

Pred. No.: 7.61e-42 Length: 1890  
Score: 417.00 Matches: 87  
Percent Similarity: 64.56% Conservative: 46  
Best Local Similarity: 42.23% Mismatches: 73  
Query Match: 15.01% Indels: 0  
DB: 4 Gaps: 0

US-09-868-987-14 (1-552) x US-09-252-991A-6330 (1-1890)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 1241 ATCATCAGCTCGGACCATTCAGTCGCGCTGGGCAACAGTTCGGTATACCCGCTC 1300  
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 1301 GATGCGCGCGGAGTCTCCAGCTGCGCTGCTGCGCGCGGCGCTGCTGCG 1360  
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 1361 CCGATGTACTTCGCGCAAGAACGACCATTCGCGCGAGCTGGGCGCGGCAACATCGCC 1420  
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 1421 AAGGCTATCGATCGCTCTGCGGCATGCTCTTCGTATCGCTTTCATCATCGTCATC 1480  
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeuLeuIleTyr 100  
Db 1481 TACCGTTTCTCGCGGTGATCGCCACCTGCTGCTGCTGCTGCTTCAACATGTCATGCTGTG 1540  
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1541 GCGCTGATGTCGATCTTCGCGCGACCTTGACCTTCGCGCGCATCGCGGTATCGTGTG 1600  
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
Db 1601 ACCATGGCATGGCGGTGGACGCCAAACGCTGTGATCTTCTCGCGGATACGCGAGAACTG 1660  
Qy 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
Db 1661 GCCAATGGCATGTCGCTCCAAAGCGCATCCACGAAGGCTTCAACCGGCTTCAACCGCG 1720  
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Db 1721 ATCTCTGACGCGCAACCTGACCTGCTGCTGCTGCGCGCATCTTATACGCCATGGGCA 1780  
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Db 1781 GGCCCGGTGAAGGATTTCGCGTACCATGTCGCTGCGCATCATCACCTCGATGTTTACG 1840  
Qy 201 AlaLeuPheMetThrLys 206  
Db 1841 GCCATCATGGTACCCCGC 1858

## RESULT 13

US-09-252-991A-6245

Qy	444	AlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAenThrLeuIleIlePhe	463
Db	604	GTCTCGCCGCTGTGTCGGCGTGGTGTACTCGCTGAACGACACCATCGTGTATCTTC	663
Qy	464	AspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHis	479
Db	664	GACCGGTGGCGGAGAACTTCGCGGTGCTCGGCAAGGCCGATCTGGTCGAGAACCTGAAC	723
Qy	480	ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr	499
Db	724	ATC-----TCCACACCGACAGACCTGTGTGGTACCATCGCCACCTCGGTTC	771
Qy	500	ThrLeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPheAenPheAla	519
Db	772	ACGTCGTGCTGATCGCGCGCTGCTGTTCTTCGGCGGCACACCTGTTTCGGTTCTCC	831
Qy	520	PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuYrIleAlaProPro	539
Db	832	ATCGCCCTGTTCGTCGGTGTAAATGGCGGTACCTACTCTGTCGATCTACATCGCCAACGTG	891
Qy	540	LeuLeuLeuPheMet	544
Db	892	GTATTGATCGGCTG	906
RESULT 14			
US-09-252-991A-6054/c			
; Sequence 6054, Application US/09252991A			
; Patent No. 6551795			
; GENERAL INFORMATION:			
; APPLICANT: Marc J. Rubenfield et al.			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD			
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 107196.136			
; CURRENT APPLICATION NUMBER: US/09/252.991A			
; CURRENT FILING DATE: 1999-02-18			
; PRIOR APPLICATION NUMBER: US 60/074,788			
; PRIOR FILING DATE: 1998-02-18			
; PRIOR APPLICATION NUMBER: US 60/094,190			
; PRIOR FILING DATE: 1998-07-27			
; NUMBER OF SEQ ID NOS: 33142			
; SEQ ID NO 6054			
; LENGTH: 984			
; TYPE: DNA			
; ORGANISM: Pseudomonas aeruginosa			
US-09-252-991A-6054			
Alignment Scores:			
Pred. No.: 2,26e-36 Length: 984			
Score: 369.50 Matches: 91			
Percent Similarity: 51.69% Conservative: 77			
Best Local Similarity: 28.00% Mismatches: 102			
Query Match: 13.30% Indels: 55			
DB: 4 Gaps: 10			
US-09-868-987-14 (1-552) x US-09-252-991A-6054 (1-984)			
Qy	227	PheValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGly	246
Db	927	TTCATGGGCATCCGCAAC-----GTGCGCTTT	901
Qy	247	SerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAenSerVal-----	264
Db	900	GCGTCACCTGATC-----CTGACCGTCATCGCCCTGGCGAGCTGTTTCACCAAGGGCATC	844
Qy	265	--LeuGlyMetAspPheLysGlyGlyTrpAlaPheThrPheAenProLysGluHisGly	283
Db	843	AACTTCGGCTCGACTTCACCGCGCGGTACGCTGATCGAGCTACGACCAACCG---787	
Qy	284	IleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGlyLeu	303
Db	786	---GCCGATCTCGCAAGGTGCGTGGTCAATGGTC-----GGCGCC	748



```

Qy 304 SerSerArgAspPheAraGlnThrPheGlySerSerGluLysIleLysIleTyrPhe 323
Db 747 GCCTACGAGGACCGCTGTGTGAGAGCTTCGGCGAGCGCGAGCTGTGTGCGGATG 688
Qy 324 SerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLysLeuThrIle 343
Db 687 CCCAGCGAGGACCGGAACTGGGCAAGAGGTGCTACGCTTCGACGAGCGCGAT--- 631
Qy 344 MetSerTyrArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrClyAsn 363
Db 630 -----GCCGGC 625
Qy 364 SerLysArgAsnAlaLysPheTrpSerLysValSerLysLysLeuSerLysLysMetArg 383
Db 624 AACCGCGGAACCTGAAGCGGTCCAGTACGTCCGCCCGCAGGTGGGTGAAGAAGTGGC 565
Qy 384 TyrGlnAlaThrIleGlyLeuLeuGlyAlaIleLeuLeuTyrValSerLeu 403
Db 564 GACCAGGCGCGCTCGGATGCTCTGGCGGTGGCGGCATCCTGTCTAGCTGGCTTC 505
Qy 404 ArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAla 423
Db 504 CGCTTCAGTGAAGTTCGCCCTCGGTGCGATCTCTCGGTGGTGACGATGCGATCATC 445
Qy 424 ThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGln 443
Db 444 GTGATGGCGGTGCTG-----TCGTTCTTC-----CAGGTGACCTTCGACCTGACC 400
Qy 444 AlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAnthrLeuIleIlePhe 463
Db 399 GTCTCTCGCGGTGTGTGCGGTGTGTCTACTCGTGAACGACACCATCGTATCTTC 340
Qy 464 AspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHis 479
Db 339 GACCGGTGCGCGAGAACTTCGCGGTGCTGCGAAGCGCGATCTGGTCGAGAACCTGAAC 280
Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThr 499
Db 279 ATC-----TCCACGACGACAGACCTCTGTGGTACCATCGCCACCTCGGTTTC 232
Qy 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnAla 519
Db 231 ACGTGTGTGCTATCGCGCGCTGTCTTCTTCGCGCGCGACAACTGTTCTGTTCTCC 172
Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db 171 ATCCGCTGTTCGTCGGTGTAAATGCGCGGTACTACTCTGCTGATCTACATCGCCCAACGTG 112
Qy 540 LeuLeuLeuPheMet 544
Db 111 GTATTGATCTGGCTG 97

```

## RESULT 15

```

US-09-252-991A-6217/c
; Sequence 6217, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6217
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6217

```

```

Alignment Scores:
Pred. No.: 1,49e-35 Length: 429
Score: 358.00 Matches: 76
Percent Similarity: 65.43% Conservatives: 30
Best Local Similarity: 46.91% Mismatches: 36
Query Match: 12.89% Indels: 20
Db: 4 Gaps: 3

```

US-09-868-987-14 (1-552) x US-09-252-991A-6217 (1-429)

```

Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuTrp 100
Db 428 TACGTTTCTTCGGCGTATCGCACCGCTTGGCTGGCTTCAACATGGTATCGTGGT 369
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 368 GCGTGTATGTCTGATCTCTCGCGCGACCTGACCTGCGCGCATCGCGGTATCGTCTG 309
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140
Db 308 ACCATGGGCAATGGGTGGAGCAACGTCTGATCTCTCGCGGATACGCGAGAACTG 249
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 248 GCCAATGGCATGTCCGTCCAAAAGGCGATCCACGAGGCTTCAACCGTCCCTTACC 189
Qy 161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 188 ATCTCGACGCCAACCTGACCTCGCTGTGTGCGCGGATCTCTATACGCCATGGGCACC 129
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Db 128 GCGCCCGGTGAAGGATTTGCGGTGACCATGTCTCGCGCATCATCACCTCGATGTTCA 69
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220
Db 68 GCCATCATGTGTCACCCGC----- 51
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db 50 ---GCAATGGTCAACCTGATCTTCGGC---GGCGGTGACTTC-----AAGAAG 9
Qy 241 LeuTrp 242
Db 8 CTGTGG 3

```

Search-completed: October 28, 2003, 18:09:44

Job time : 8948 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 422 Seconds  
(without alignments)  
3507.948 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNPLKNHASVSGK.....SLYIAPLLLFMVKENRSK 552

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO-0001/US09868987/runat\_28102003\_152603\_23869/app\_query.fasta\_1.711  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09868987@cgn\_1\_1\_271\_@runat\_28102003\_152603\_23869  
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2778	100.0	1864	9	US-09-886-468-1 Sequence 1, Appli

C	2	719	25.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
	3	678.5	24.4	2211	9	US-09-899-482-1	Sequence 1, Appli
	4	664	23.9	1584	10	US-09-974-300-1840	Sequence 1840, Ap
C	5	509	18.3	80557	15	US-10-080-170-647	Sequence 647, App
	6	506	18.2	9025608	14	US-10-156-761-1	Sequence 1, Appli
C	7	429	15.4	3309400	10	US-09-738-626-1	Sequence 1, Appli
	8	408.5	14.7	1848	14	US-10-260-877-119	Sequence 119, App
	9	393	14.1	1671	10	US-09-881-752A-95	Sequence 95, Appl
	10	350	12.6	1035	12	US-09-882-227-401	Sequence 101, App
	11	342	12.3	975	14	US-10-260-877-129	Sequence 129, App
	12	335	12.1	1298	10	US-09-974-300-6224	Sequence 6224, Ap
	13	311.5	11.2	609	8	US-08-781-986A-671	Sequence 671, App
	14	283.5	10.2	595	8	US-08-781-986A-1128	Sequence 6816, Ap
	15	275.5	9.9	1104	14	US-10-156-761-6816	Sequence 6815, Ap
	16	270	9.7	1752	14	US-10-156-761-6815	Sequence 1830, Ap
	17	259.5	9.3	1932	10	US-09-738-626-1830	Sequence 1, Appli
	18	259.5	9.3	1960	9	US-09-852-053-1	Sequence 1, Appli
	19	238.5	8.6	837	9	US-09-861-451A-27	Sequence 27, Appl
	20	228	8.2	1209	10	US-09-738-626-1828	Sequence 1828, Ap
	21	228	8.2	1562	9	US-09-852-053-2	Sequence 2, Appli
	22	177	6.4	105184	11	US-09-847-513A-1	Sequence 1, Appli
	23	169.5	6.1	2691	14	US-09-974-300-2170	Sequence 2170, Ap
	24	161.5	5.8	3159	14	US-10-156-761-6017	Sequence 6017, Ap
	25	144	5.2	3105	9	US-09-815-242-6234	Sequence 6234, Ap
	26	139.5	5.0	3168	9	US-09-815-242-8091	Sequence 8091, Ap
	27	138	5.0	2190	10	US-09-738-626-696	Sequence 696, App
	28	136	4.9	3099	9	US-09-815-242-7048	Sequence 7048, App
C	29	136	4.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
	30	135	4.9	38734	11	US-09-373-658-30	Sequence 30, Appl
	31	128.5	4.6	7888	12	US-10-085-959-28	Sequence 28, Appl
	32	128	4.6	3087	9	US-09-815-242-7419	Sequence 7419, App
	33	126.5	4.6	3123	9	US-09-912-020-122	Sequence 122, App
	34	126.5	4.6	3123	9	US-09-815-242-6106	Sequence 6106, Ap
	35	126.5	4.6	3123	12	US-10-287-274-271	Sequence 271, App
	36	136.5	4.6	15393	14	US-10-114-170-191	Sequence 191, App
	37	125.5	4.5	3078	12	US-09-912-020-123	Sequence 123, App
	38	125.5	4.5	3078	12	US-10-287-274-272	Sequence 272, App
C	39	125.5	4.5	19061	12	US-10-085-959-148	Sequence 148, App
C	40	125.5	4.5	19061	12	US-10-085-959-254	Sequence 254, App
C	41	125.5	4.5	48254	12	US-10-238-075-829	Sequence 829, App
	42	123.5	4.4	2271	14	US-10-156-761-2139	Sequence 2129, App
	43	123	4.4	3179	12	US-09-882-227-187	Sequence 187, App
	44	123	4.4	3280	8	US-08-834-705-17	Sequence 17, Appl
	45	122.5	4.4	3189	9	US-09-815-242-4056	Sequence 4056, Ap

ALIGNMENTS

RESULT 1

US-09-886-468-1  
; Sequence 1, Application US/09886468  
; Patent No. US20020037293A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis Pasteur Limited  
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there  
; FILE REFERENCE: 77813-5  
; CURRENT APPLICATION NUMBER: US/09/886,468  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/113,280  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,281  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,282  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,283  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,284  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,285  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,385  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/114,050

```
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1756)
US-09-868-468-1

Alignment Scores:
Pred. No.: 4,698-304 Length: 1864
Score: 2778.00 Matches: 552
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-868-987-14 (1-552) x US-09-868-468-1 (1-1864)

QY 1 MetValSerSerProIleLeuAenValProLeuLysAsnHisAlaSerValSerGlyLys 20
DB 101 ATGCTCAGCAGCCCTATTTTAAACGTCCTCCATCGCAATCATGCGAGTGTCTCAGGAAA 160
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
DB 161 TTTACCCACCGTGAGTGAGCAAACTCGCTCAGATTTAAATCTGGAGCGATGTCCTTT 220
QY 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
DB 221 GTTCCCGAGGTTCTCAGTGAAGACAGCATCTCTTCATCTTGGAAAAAACAATGTACA 280
QY 61 GlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 80
DB 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTGTATGAGCGTATAT 340
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuAsnLeuLeuIleTrp 100
DB 341 TATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTCGAAATCTTTTGTCTTATCTGG 400
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
DB 401 GCAGCTCTACAGTATTGGATGGCCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTT 460
QY 121 AlaMetGlyMetAlaValAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
DB 461 GCTATGGGGATGGCCGTPAGATGCAAAATGTTCTTGTATTGAAAGAAATCCGAGAGGAAAT 520
QY 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
DB 521 TTATTGTCTCAAGTCTTAAATAATCTGTAGAAAAGAGATATACCAAGGCTTTTGGAGCC 580
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThr 180
DB 581 ATTTTTCATTCTAATCTGACTACAGTATTGGCCTCAGCATCTCTTTCTTCTCTAGATACA 640
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPheThr 200
DB 641 GGGCCTATTAAAGGGTTGCTTTTGACATTGATTTAGGAATTTTCTCTCAATGTTTACG 700
QY 201-AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220
```

RESULT 2

US-10-329-960-1/c

```
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, .Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```



```

Db 267448 TGTGGTGAAGAAATAGCACAGGTGCGGTATATATCGCATTTAGCGACATTAGCAATGGT 267389
Qy 397 eLeuLeuThrValSerLeuArgPheGluTrpGlnThrAlaPheSerAlaValCysAlaLe 417
Db 267388 GCTTATTTATGTGGGTGCACGTTTTGAATGGGTTTAGCGTTTGGCAGTATCGCTTCTCT 267329
Qy 417 uileHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLY 437
Db 267328 TCGGCACACGTCATTATTAGCTAGGGGTA-----TTCTCTGCATTACA 267284
Qy 437 sileGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAs 457
Db 267283 AATTGAAATTCATCTTACTTTGTGCGCAGCATTTATCTGTGGTGGTACTCCATCAA 267224
Qy 457 pAsnThrLeuIleLeuPheAspArgIleArgGluAsp---ArgGlnAlaAsnLeuPheTh 476
Db 267223 CGATAGTATTGGGTATTGACCGGGTTCGTGAAATTTCCGAAAAATTAGACGATTGGA 267164
Qy 476 rProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetTh 496
Db 267163 TACGATTGCATATT---ATTGATATTCTTTAACGCAAACTTTATCAAGAACTATCATTAC 267107
Qy 496 rThAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPh 516
Db 267106 TTCGGTTACTACATTAGTTGTCGTATGCGCATTTGTTCTTTGGTGGTCTCTCCATTC 267047
Qy 516 eAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrI1 536
Db 267046 TAACTTTTCACTTGTCTTACTCGTAGGTATTGGATTGTTGTTACTTATTCCTCGATTGTT 266987
Qy 536 eAla 537
Db 266986 TGCC 266983

```

## RESULT 3

```

US-09-899-482-1
; Sequence 1, Application US/098999482
; Patent No. US2002006641A1
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
; FILE REFERENCE: GC385-PCT
; CURRENT APPLICATION NUMBER: US/09/899,482
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Bacillus Subtilis
US-09-899-482-1

```

## Alignment Scores:

```

Pred. No.: 3.73e-66 Length: 2211
Score: 678.50 Matches: 172
Percent Similarity: 50.25% Conservative: 126
Best Local Similarity: 29.01% Mismatches: 198
Query Match: 24.42% Indels: 98
DB: 9 Gaps: 15

```

US-09-868-987-14 (1-552) x US-09-899-482-1 (1-2211)

```

Qy 2 ValSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGlyLys 20
Db 589 GTATCCGCTCCAAATGTAAAGTCAGGAACCTAAATACAACTGATGTAATAAATTAAGAGGTGAT 648
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyValAlaMetSerPhe 40

```

```

Db 649 TTCACAGCTCAAGAAGCGAAAGATTATAGCCAGCATTTTAAACGCGAGCGCACTT----- 702
Qy 41 ValProGluValLeuSerGluGlu-----ThrIleSerSerAspLeuGlyLys 57
Db 703 ---CCTGTGAACACTGACTGAAAGATATTTCGACATCATAGTCGGCGCAATTCGGCCAGCAG 759
Qy 58 GlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMet 77
Db 760 GCTCTCCATGATACCGTGTGTTCCGGTATTGTCGGTATCGCAATTAATTTCTTATTATG 819
Qy 78 SerValTyrTyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeu 97
Db 820 CTTTCTTATTACCGTCTCGCGGGAATATC-----GCGGTGATTACGCTGCTGTT 870
Qy 98 LeuIleTrpAlaAlaLeuGln-----TyrLeuAspAlaProLeuThrLeuSerGly 114
Db 871 TATATCTACATTACCTCCAGATCTTTGACTGGATGAATGCCGTACTCACGCTTCCGGGA 930
Qy 115 LeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGlu 134
Db 931 ATTGCCGCTCTCATTTTAGGTGTCGGGATGCTGTTACGCCAACATTAATACCTATGAG 990
Qy 135 ArgIleArgGluGluPheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyr 154
Db 991 CGGATTAAGAAGAGCTCAAGCTAGGAAGTCAGTCGCTCTGCTTCCGTTCCAGAAAC 1050
Qy 155 ThrLysIlePheGlyAlaIlePheAspSerAsnLeuThrValLeuAlaSerAlaLeu 174
Db 1051 AGACGGTCATTTCGACGATTTTTCGCGCAATATTACACCATTAATTTGACGCGGTGTG 1110
Qy 175 LeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIle 194
Db 1111 CTCTTATCTTTGGGCAAGCTCTGTTAAAGGTTTCGACAACTGCTGATCTATCGATT 1170
Qy 195 PheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMetLeu----- 211
Db 1171 TTGACAAAGCTTATACACTGCGGTTTCTTATCGAGATTTCTCCTCGCTCTCTTGTGAA 1230
Qy 212 -----TrpMetAsnLys----- 215
Db 1231 AGCAGATGGCTTGATCGGAAAAAGGCTGGTGTGTCATAAGAACATATCATGGAT 1290
Qy 216 -----ThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIleLys 231
Db 1291 ATTCAGGATACGGATGAAATACAGACCCGATACGCCATTCCTCAA-----AAA 1338
Qy 232 HisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeu 251
Db 1339 TGGGATTTACAGAGCAAAACGCAATACTTCTTTATTTCTCCAGTCGGGTACGCTTGC 1398
Qy 252 GlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGly 271
Db 1399 GGGATTAATTATCTGCTGTTGTTTACGCTGAAT-----CTTGGCATTTGACTTTGCAAGC 1452
Qy 272 Gly-----TyrAlaPheThrPheAsnProLysGluHisGly 283
Db 1453 GGTGCACGGATTGAGTGCAAAGCGCACCATAGCTGACGACGAGCAAGTTTGACAGGAT 1512
Qy 284 IleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGlyLeu 303
Db 1513 TTTGAATCTCTGGGTATGGACCTGATCTAGTGTCTGTCAGCGCAAAAGAGCAATATC 1572
Qy 304 Ser---SerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyr 322
Db 1573 GGTGTTGCCGTTTGTGCGGGTGCAGATAAAACCATTTGCCAAAGTAAACAGTAT 1632
Qy 323 PheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr 342
Db 1633 TTTAAAGACAAA----- 1644
Qy 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362
Db 1645 -----TACGGA 1650

```

```

QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382
D 1651 TCTGATCAAAATGACACACAGTTTCACCGACAGTCGTGAGGAGCTGGCGAGAAATGCG 1710
QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSer 402
D 1711 CTGTACGAGCTGTCTATAGCTTCTATTGGC-----ATCATATTACGTTTCA 1758
QY 403 LeuArgPheGlnTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422
D 1759 ATCCGATTGGAATACAAAATGCGGATTCCTGCCATCGCTCATTCGCTATATGAC----- 1812
QY 423 AlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeu 442
D 1813 -----GCATTCTTTATCGTCACGCTTCCTAGTATTACAGGCTTGAGGTAGATGTT 1863
QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIle 462
D 1864 ACATTATCGCGGCATCTTGACGATAATCGGTTATTCATTACGATACATCGTTTACA 1923
QY 463 PheAspArgIleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHis 479
D 1924 TTTGACAGGTCGCGGAGCATATGAAAAGCGTAAGCCGAAACCTTTGCCGATCTGAAC 1983
QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThr 499
D 1984 CATATTGTAACCTGAGCCCTGCAGCAACCTTTACACGTTTCATTAACACATGTTTAAAC 2043
QY 500 ThrLeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAla 519
D 2044 GTTGTGATTGTTGTGACATCTGCTGATCTTTGGAGCATCTTCTATTACTTCTCA 2103
QY 520 PheIleMetThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
D 2104 ATTGCTTTATGTCGGCGCTTAAACAGCGCTTTATCTCTCTATACATTGCGGCACAA 2163
QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
D 2164 ATTTGGCTTGC-ATGGAAGAAGAGAGAACTGAAAAAAGA 2201

RESULT 4
US-09-974-300-1840
; Sequence 1840, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; CURRENT APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1840
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1840

```

## Alignment Scores:

```

Pred. No.: 9,52e-65 Length: 1584
Score: 664.00 Matches: 173
Percent Similarity: 48.98% Conservative: 114
Best Local Similarity: 29.52% Mismatches: 177
Query Match: 23.90% Indels: 122
DB: 10 Gaps: 16

```

US-09-868-987-14 (1-552) x US-09-974-300-1840 (1-1584)

```

QY 17 ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGly 36
D 21 ATTAAGGCAACTTTACGATTACGAAAGCAAAAGACATGCGAGCATTTTAAATGCGGA 80
QY 37 AlaMetSerPheValProGluValLeuSerGluGlu-----ThrIleSerSerAsp 53
D 81 GCTCTT-----CCTGTCAAACTCGTTGAAAGATATTCGACATCCCGTCGAGCGCAG 131
QY 54 LeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeu 73
D 132 TTTGGGACGACGCTTGAATGATACCGTGTTCGCGGAATCATTCGTATGCTATTATT 191
QY 74 IleValLeuMetSerValTyrTyrArgPheGlyValIleAlaSerGlyAlaValLeu 93
D 192 TTTCTATTATGCTTCTTACTATCGCTTGCCTCGCTTATT-----GCGGTCAAT 242
QY 94 LeuAsnLeuLeuIleTrpAlaLeuGln-----TyrLeuAspAlaProLeu 110
D 243 AGCTTTTCGCTTACATCTACATTACATTCACGTCGAGATATTCGACTGGAGTATCGGTTCTG 302
QY 111 ThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnVal 130
D 303 ACCCTTCGGGATCGGCTCTTATATTGGCGTCGGATGCGGTTTCACGCCAACATC 362
QY 131 LeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLysSerVal 150
D 363 ATTACGTATGACCAAGCAAGAGCTGGAAGCTCGGCAAAATCGGTCTCGGCTTT 422
QY 151 GluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThrValLeu 170
D 423 AAAGCAGGAAACAGAGATCGTTTGCACAGATTTTGTATGCCAATATACGACGATGCTT 482
QY 171 AlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu 190
D 483 GCGGAATCGCTTTTCATATTGTTACAACTCTGTCAAGGCTTCGCGACCATGCTG 542
QY 191 IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMet 210
D 543 ATCTATCATCTTTGACAAAGCTTTATTACAGCTGTCTTCCTCGCGCTTCCTGCTCGGC 602
QY 211 Leu-----TrpMetAsnLysThrGln----- 217
D 603 CTGCTGTCGAAAGCGCTGCTTCACCGAAAAAGAGCTGTTTGGGTGCGCAAAAAA 662
QY 218 -----HisThrGlnLeuHisMetMetAsnLysPheValGly 229
D 663 GACATTTTGGATATCAGAAAGACAGACGAAATACGGAGCCTCCGAAACCGTTT----- 716
QY 230 IleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPhe 249
D 717 TCCAAATGGGATTCGTCGGGAAGCGCAAAATGTTCTTGTGCATTTTCGGCGCTTCTCTG 776
QY 250 LeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuMetAspPhe 269
D 777 GCGCTGCTTATCGTCTTGTGTTAAGCTGAAC-----CTCGCATCGATTTC 830
QY 270 LysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGln 289
D 831 TCAAGCGGA-----TCGACA 845
QY 290 MetArgGlyLysValValHisLysLeu-----GlnGlu 300
D 846 ATCGAAGTTTCAAAGCGATCATAAGCTGACACCCAGCAGCTCGAAAAAGGACTTTGAACAA 905
QY 301 AlaGlyLeu-----SerSerArgAspPheArgIleGln 311
D 906 GTCCGGCTTGATCTCGATCGTCTTTCAGGCGCAAAAAATGATCATCGGCTCGCC 965
QY 312 ThrPhe-----GlySerSerGluLysIleLysIleTyrPheSerAsp 325
D 966 AGGTTTGTGCGGCTGCCGAATCAGAAAAAGATCGCTGAAGTCAAGATTTATTTAAAGAC 1025

```

Qy 326 LysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSer 345  
 |||  
 Db 1026 AAA----- 1028  
 Qy 346 TrpArgTyrCysGlyIleValIleValArgAsnArgProArgPheLeuTyrGlyAsnSerLys 365  
 |||  
 Db 1029 -----TACGGAAGCGAACCA 1043  
 Qy 366 ArgAsnAlaLysPheTrpSerLysValSerSerLysLysLeuSerLysLysMetArgTyrGln 385  
 |||  
 Db 1044 AATGTCAGCACCGGTATCGCTTACCGTCGGCAAGAGCTGGCGAGAAATGCGCTTATATGCG 1103  
 Qy 386 AlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPhe 405  
 |||  
 Db 1104 GTGATCATCGCTCAATCGG-----ATCATTTGTATGCTCGATCGGCTTT 1151  
 Qy 406 GluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCys 425  
 |||  
 Db 1152 GAATACAAGATGGCGATTGCGCTTACTTCAATTGCTGTATGAC----- 1196  
 Qy 426 AlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeuGlnAlaIle 445  
 |||  
 Db 1197 GCGTTTTTCATGTGCGCTTTCAGCGCTGACAAAGGCTTGAGGTGATGTTTACTTTTCATT 1256  
 Qy 446 GlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArg 465  
 |||  
 Db 1257 GCTCCGCTGTCAGCATCATCGGTATTTCGATAATGATACATGTAACGTTTGACAGG 1316  
 Qy 466 IleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuVal 482  
 |||  
 Db 1317 ATCAGGGAGCATATGAAAAAGCGAAACCGAAACCTTCAGCGACTTGTGCGCATATTGTG 1376  
 Qy 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502  
 |||  
 Db 1377 AATTGAGCTTTCAGCAGACGCTTCACGCGTTTCGATTAAATACGGTGTGACTGTGCTCAT 1436  
 Qy 503 ValLeuLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAlaPheIleMet 522  
 |||  
 Db 1437 GTGCTCATGTCATCTCTCATCTTCGGCGCTGCATCGATTTCGAACTTCTCTGTTCGCGTG 1496  
 Qy 523 ThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu 542  
 |||  
 Db 1497 TTGTCGAGCTTTTGAGCGGAGTATATTCACTCACTTACATCGCAGCTCAACTATGCTT 1556  
 Qy 543 PheMetValArgLysGlu 548  
 |||  
 Db 1557 GTCTGGAAGGCGCGAA 1574

## RESULT 5

US-10-080-170-647/c  
 ; Sequence 647, Application US/10080170  
 ; Publication No. US20030129601A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: COLE, S.T.  
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
 ; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
 ; TREATMENT OF MYCOBACTERIOSES  
 ; FILE REFERENCE: 03495.0218  
 ; CURRENT APPLICATION NUMBER: US/10/080.170  
 ; CURRENT FILING DATE: 2002-06-10  
 ; PRIOR APPLICATION NUMBER: 60/270,123  
 ; PRIOR FILING DATE: 2001-02-22  
 ; NUMBER OF SEQ ID NOS: 652  
 ; SOFTWARE: Patent in ver. 2.1  
 ; SEQ ID NO 647  
 ; LENGTH: 80557  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; US-10-080-170-647  
 Alignment Scores:  
 Pred. No.: 2.02e-44 Length: 80557

Score: 509.00 Matches: 160  
 Percent Similarity: 42.58% Conservative: 124  
 Best Local Similarity: 23.99% Mismatches: 221  
 Query Match: 18.32% Indels: 162  
 DB: 15 Gaps: 18  
 US-09-868-987-14 (1-552) x US-10-080-170-647 (1-80557)  
 Qy 2 ValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys--- 20  
 |||  
 Db 66923 CTGACTCGCAGGTGTGTTCAGTCACCGCAGATCCAGGAACGATCCCGCGCGCGCAACC 66864  
 Qy 21 -----PheThrHisArgGluValSerLysLeuAlaSerAsp 32  
 |||  
 Db 66863 CAGATCAGCGGTGGGATCCGCATTCACCGCGGACCGCAGCTCGCAACGTC 66804  
 Qy 33 LeuLysSerGlyAlaMetSerPheValProGluValLeuSerGluGluThrIleSerSer 52  
 |||  
 Db 66803 CTCAAAGTACGGGTGCTGCTCGCTGCTCTTCGAAACCGTCGGAAGCCCAACGGTTTCGCG 66744  
 Qy 53 AspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72  
 |||  
 Db 66743 ACCTTGGGTTGTCATCGCTGCGCGGCGCATGATCGGGTGGATCGGCTGCTGTTG 66684  
 Qy 73 LeuIleValLeuMetSerValTyrArgPheGlyGly-----Val 86  
 |||  
 Db 66683 GTGCTGTGTATTGCTGTCTTACTACCGAGTGTGGACTGTCTCAGCGCTTGTGCGTA 66624  
 Qy 87 IleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaLeuGlnTyrLeu 106  
 |||  
 Db 66623 GTGCTTCTGCTCAATGTTTTCGATCTGCTGCTTGGC-----CGATATATC 66570  
 Qy 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaVal 126  
 |||  
 Db 66569 AACTACACCTAGATCGCTGTGTTTTCGGGCTGATATCGGTATCGGACCAACCGCC 66510  
 Qy 127 AspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeu 146  
 |||  
 Db 66509 GACTCGTTCGTGTGTTCTTCGAACGATCAAGACGAGATCCGGAAGCGCTTCTGTT 66450  
 Qy 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeu 166  
 |||  
 Db 66449 CGGTGCGCAGTGGCGCGTGGCGCGTCCCGCAAGACGATCGTGTGGGCAACGCC 66390  
 Qy 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186  
 |||  
 Db 66389 GTCACTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 66330  
 Qy 187 AlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLys 206  
 |||  
 Db 66329 GCGTTCACCTG-----GGCTCACACGATCCTTGACCTCGTCTG 66288  
 Qy 207 PhePhePheMetLeuTrp-----MetAsnLys 215  
 |||  
 Db 66287 GTGTTCTGTGTGACGTGGCGTGTGTGTATCTGCGGTCCAAATCGTCTGCTGGCCAAG 66228  
 Qy 216 ThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyLys----- 231  
 |||  
 Db 66227 CCGGCATACAAACGCTGGGAGCGGTGCGACAGGTGCGACGAAACCGCGGCGCATGGCT 66168  
 Qy 231 ----- 231  
 Db 66167 CGGACGGGAGGATAGCGATGGCGTCGAAAGCAAGAACCGCGCGGAGCAAGGCC 66108  
 Qy 231 ----- 231  
 Db 66107 ACGTGCGCGGTGGAATCACCAGGCGCACGAGAGTCTGTGGCGGAGCCGACGGCGAC 66048  
 Qy 232 -----HisAspPheLeuArg----- 236  
 |||  
 Db 66047 AGCACCAACAGACCGCTCGAAGCTGGGCGACCATAGCTTCCTTTTCGCGCTGTACACC 65988  
 Qy 237 -----GlyCysLysLysLeuTrp---AlaValSerGlySer 247

```
Db 65987 GGCACCGGGCGTTGAGTAGTGGAGCGCGCGTGTGTGGTGGCGTGCGGCGG 65928
Qy 248 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet 267
Db 65927 ATCGTCGGGTGGCCATCGCCAGCATCGTCTTTCGTGGTTC-----ACCTTCGGCATC 65874
Qy 268 AspPheLysGlyGlyTyAlaPheThrPheAsnProLysGluHisGlyIleSerAspVal 287
Db 65873 GATTTCAAAGCGGTACACCGGTGTGTTTC-----CCGCGTGGCAGCACCAAGTC 65823
Qy 288 AlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGlyLeuSerSerArgAsp 307
Db 65822 GCGCAGGTGCAAGACGCTACTACCGGCCCTC-----GGCAGTGAGCCCACTCG 65772
Qy 308 PheArgIleGlnThrPheGlySerSerGlnLysIleLysIleTyPheSerAspLysAla 327
Db 65771 GTGGTGATCGTCGGGGCGGTGCTTCGGCAGCGGTTCGAGATTCGTCGAGACCTGACC 65712
Qy 328 LeuSerTyThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArg 347
Db 65711 AGCCACGACGCGCAAGCTCGGACGCCCTATTTCGAGGCCCTTC----- 65667
Qy 348 TyrCysGlyIleValValArgAsnArgProArgPheLeuTyArgAsnSerLysArgAsn 367
Db 65666 -----GGGCCCAAGGAACCGACCGTGCAGATCACCAAGAGCGCGTG 65631
Qy 368 AlaLysPheTrpSerLysValSerLysLysLeuSerLysLysMetArgTyGlnAlaThr 387
Db 65630 GCCATCAGCGACTCGGGGTGTTCGGAGACCTGGGGCGGTGCAGATCACCAAGAGCGCGTG 65571
Qy 388 IleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyValSerLeuArgPheGluTrp 407
Db 65570 ATCCGCTGGGTGGTTCGTGGTCTGTCGCCCTTACATTACCGTGGCGTACGACGCG 65511
Qy 408 GlnTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaVal 427
Db 65510 TACATGACCATCTCGCGGATAACGGCATGCTCTTCGACCTGACCGTCACCGCGCGGTG 65451
Qy 428 LeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAla 447
Db 65450 TATTCCCTGTGGGTTC-----GAAGTCACCCCGCGCACGGTCATCGGC 65406
Qy 448 LeuMetThrValLeuGlyTySerLeuAsnAsnThrLeuIlePheAspArgIleArg 467
Db 65405 CTGCTGACCATCTTCGGGTCTCGCTCTATGACACCGTCTATGTTTCGACAGGTTCGAG 65346
Qy 468 GluAsp-----ArgGlnAlaAsnLeu 474
Db 65345 GAGAACCCATGCTCCAGCACACACCGCGGCACCTTCGCGGAGAGGCCAACCTG 65286
Qy 475 PheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVal 494
Db 65285 -----GCGATCAACCAAGAGCTTCATCGCGTGCATC 65256
Qy 495 MetThrThrAlaThr---ThrLeuSerValLeuMetLeu-----Phe 509
Db 65255 AACCCAGCTGATCGCGGTGTTCGCGGTGTGGCGCTGATGGTGGCGGTCTGGCTG 65196
Qy 510 IleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGly 529
Db 65195 CTGGCGCTCGGCACCTCTCAAGACCTCGCGCTGGTGCAGCTGATCGGCATCATCGGC 65136
Qy 530 ThrLeuSerSerLeuTyIleAlaProProLeuLeuLeuPhe----- 543
Db 65135 ACCTACTCGTCAATCTCTTCGCGACCGCGTGTGTGTATACCTCCGTCGTCGACGCGAG 65076
Qy 544 MetValArgLysGluAsnArg 550
Db 65075 TTGGTGGCGCAACACACCCCGT 65055
RESULT 6
US-10-156-761-1
```



```
Db 8161298. CGTCCCGCGGACCAATCCTGGTCTCCGACTTCGTGTCTCTCGCCGCGCGGTGCTG 8161357
Qy 176 PhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu----- 190
Db 8161358 TTTCATCGTGACCGTCGGCAAGCTCCAGGCTTCGGGTTACGGTCTGACACACCGTG 8161417
Qy 191 -----IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLys 206
Db 8161418 CTCGACGTGGTGGTCTGCTCTCTTCCACCAAGCGCTGATGACGATCTCGCCGCGCAAG 8161477
Qy 207 PhePhePheMet-----LeuTrpMetAsn-LysThrGlnHisThrGlnLeu----- 221
Db 8161478 CAGTTCTTCGGCAACGGTCACAGCTGGTCCGGCCTCGACCCGCAAGCGCTCGGCCAG 8161537
Qy 222 -----HisMetMe 224
Db 8161538 CGCCATTTCGGCGCACCGGAGTCTCTCCGCCCGCGGACCCGCAAGGAGCGTGAGAT 8161597
Qy 224 tAsnLysPhe-----ValGly-----IleLysHisAspPheLe 235
Db 8161598 GTCGAAGCTCGCAACCTCGCGCGCGGCTCCACCGAGCGAGGTGCGCTACGACTTCGT 8161657
Qy 235 uArgGlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAl 255
Db 8161658 C---GGCAACCGCAAGATCTGG---TACGGCATCTCGATCTCGATCACCATCACGGCCAT 8161711
Qy 255 aLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPh 275
Db 8161712 CTTGGCTCGAGGTGCGCGGCTGAACATGGCATCGAGTTCAGGGCGGAGCGGTCTT 8161771
Qy 275 eThrPheAsnProLys-----GluHisGlyIleSerAspValAl 288
Db 8161772 CACC---ACCCCAAGACAGCGTCTCGTGTCCGAGCCGAGAGGTACGCGGAGGAGGC 8161828
Qy 288 aGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPh 308
Db 8161829 GTCCGGCCACACGCGGATCGTGCAGAGCTCGGTCTCGCGCGGTCTGCGGTCCAGATCGC 8161888
Qy 308 eArgIleGlnThrPheGlySerSerGluLysIleLysIleTyPheSerAspLysAlaLe 328
Db 8161889 CGGTGTGACACC---GGCAAGTCGACCATCAAGCCGAGCTCTCCAGAACCTGAA 8161945
Qy 328 uSerTyThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTy 348
Db 8161946 GGTGACTCGGAAGAATCAACGCGGACCTGTGTGCG---CCAGCTCG--- 8161991
Qy 348 rCysGlyIleValArgAsnArgProArgPheLeuTyArgLysSerLysArgAsnAl 368
Db 8161992 ---GGTGAGCAGATCGCAAC--- 8162009
Qy 368 aLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyGlnAlaThrIl 388
Db 8162010 -AAGCCCTGGCAA----- 8162021
Qy 388 eGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyValSerLeuArgPheGluTrpGl 408
Db 8162022 -GGCCTGTGTATCTTCATGATCCTGGTGTGATCTACCTGCGGATCGGCTTCGATGGCG 8162080
Qy 408 nTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLe 428
Db 8162081 GATGCGGTGCGGCGCTGCTGCTGATCCAGCATCACCATCAGGTGCGGATCTA 8162140
Qy 428 uPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLe 448
Db 8162141 CGCCCTCTCGCGCTTC-----GAGGTGAGCCCGCGGTACCGTCACTCGGCT 8162185
Qy 448 uMetThrValLeuGlyTyTrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGl 468
Db 8162186 GCTCAGCATCTCGGTACTCGCTATGACACGGTCTGCTCTCGACACCCCTCAAGGA 8162245
Qy 468 u-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAl 485
Db 8162246 GCAGACGAAGGACCTTCACCAAGCAGACTCGCTGGACCTTACAGCATGTGCGCAACCGCTC 8162305

Qy 485 aLeuGlnLysThrPheSerArgThrValMetThrAlaThrThrLeuSerValLeuLe 505
Db 8162306 GATCAACAGCACCTCTGGTCCGCTCCATCAACACACCGGTGGTGGCTTCTGCGGTCGC 8162365
Qy 505 uMetLeuLeuPheIleGlyGly-----SerSerValPheAsnPheAlaPh 520
Db 8162366 GGGCTCTGCTTTCATCGCGGTGGCGTCTCGCGCGGCGACGCTCAACGACATCTCGCT 8162425
Qy 520 eIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyTrIleAlaProPole 540
Db 8162426 GTGCTCTTGTGGTCTCGCGCGGCGGTACTCTCTCATCTTCATCGCCACGCGCT 8162485
Qy 540 uLeuLeuPheMetValArgLysGlnAsnArgSerLys 552
Db 8162486 CGTCCCGCAGCTCAAGAGCGCGGAGCGCGATGAAG 8162522

RESULT 7
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738:626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 9,73e-33 Length: 3309400
Score: 429.00 Matches: 142
Percent Similarity: 42.75% Conservative: 147
Best Local Similarity: 21.01% Mismatches: 222
Query Match: 15.44% Indels: 166
DB: 10 Gaps: 18

US-09-868-987-14 (1-552) x US-09-738-626-1 (1-3309400)
Qy 1 MetValSerSerProfileLeuAsn-----ValProLeuLysAsnHisAlaSerValSer 18
Db 1759690 GTGATTTCTGCACCCCGGTATTAGTCAGCAACCCCTGGTGGTTCGCAACATCCATCACC 1759631
Qy 19 GlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAla--- 37
Db 1759630 GGTGACTTCACTCAAGCTGAAGCCCAAGATCTGCGCAACCACTGCGTACGGTGCATTG 1759571
Qy 38 ---MetSerPheValProGlu-----valLeuSerGluGluThrIleSerSer 52
Db 1759570 CCCCTGAGCTTCGAGGTGAAGAACGCGGAGCGCGGAGAACCTACCAACCGTTCCGCCA 1759511
Qy 53 AspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72
```

```
Db 1759510 TCACTAGCGCAGCATCCTTGAAGCGCGACTGATCGAGGCATCGTCCGCGCATCGCGCTG 1759451
Qy 73 LeuileValLeuMetSerValTyrArgPheGlyGlyValile----- 87
Db 1759450 GTCGCCATTCGTGTTGTCGCTACTACCGGTCTTCGGATTGTTCCCTGTTCCACCGTG 1759391
Qy 88 ---AlaSerGlyAlaValLeuLeuAsnLeuLeuLeuileTTPAlaAlaLeuGlnTyrLeu 106
Db 1759390 TTTGCCGAGCGGTGTTGGTCTACGGCTTCTGTTACTGTTGGA-----CGCTGGATC 1759337
Qy 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyileValLeuAlaMetGlyMetAlaVal 126
Db 1759336 GGATATTCCCTAGACCTTGTGGTATCGCGGTGTTGATCATCGGTATCGGTATCCGACCGCC 1759277
Qy 127 AspAlaSerValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeu 146
Db 1759276 GACTCCTTCGTGGTGTCTATGAGCGCATCAAGATGAGATCCGTGAAGGAAGATCCTTT 1759217
Qy 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeu 166
Db 1759216 AGATCTCGATCTCTGTCGATGGAAAGCGCCAGCGCACCATCTGTCAGGCAACATG 1759157
Qy 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186
Db 1759156 GTCACTTTGCTCGCGCTATCGTCAATTACTTGTCTCGGTGCGGGAAGTCAAGGGCTTT 1759097
Qy 187 AlaLeuThrLeu---IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMet-Th 205
Db 1759096 GCCTTCACCTGGGTCTGACACCGCATTCGATCTCGTCTGTGACC-----TTCTGATC 1759043
Qy 205 rLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGlnLeuHis----- 222
Db 1759042 ACGGCACCACTGGTTATCTTGGCATCAGCAACCCATCTTTGCCAAGTATCGGTCAAC 1758983
Qy 222 -----MetMetAsnLysPheValGlyle----- 230
Db 1758982 GGATGGGACGATGATGAAGCTGTTGAAGAACCGCGCCCAACGGTGAATTGGATGAG 1758923
Qy 223 ----- 230
Db 1758922 CCTGAGTACCTGAAAAAGATCCATGCAAGAAATGCGCAGCTGATAGAGCTTCCACTGAC 1758863
Qy 230 ----- 230
Db 1758862 AATCTTCCACTGACATTTCTGAAGCACCTTGGACCGATACGACCAAGAGGAGGAG 1758803
Qy 230 ----- 230
Db 1758802 TAGCATGACTGATTCCAGACTGAATCACTGTCACTCAGAGCGGTAAACCAGCCAAAA 1758743
Qy 231 -----LysHis-AspPheLeuArgGlyC 238
Db 1758742 AAGCAGTGTGTTCAACAGCCTCTACACGGGTGACGCGCATGACTTCACTCCCAAAA 1758683
Qy 238 yLysLysLeuTyrAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyP 258
Db 1758682 CCAAACTGTGTACTGATCACCAGCATTTTGTGGTTATCTCGATCTCTCATGCCCA 1758623
Qy 258 heGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyIleTyrAlaPheThrPheA 278
Db 1758622 TCCGTGGCTTCTCC-----CTGAGCATCGATTTCAGGCGGTACCAAGATGAGCATGC 1758569
Qy 278 snProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysL 298
Db 1758568 CAGCATCGGATTTACTCCACCGAACAGGTGGAGGAACC-----TTTA 1758527
Qy 298 euGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluL 318
Db 1758526 CTGAGCCACCGCATTTACTCCGAAATCGTGAGATCGTGGTTCGGCGACGCCGCA 1758467
Qy 318 yslLysIleTyrPhe-SerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSer 337
```

## RESULT 8

```
US-10-260-877-119
; Sequence 119, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: H. influenzae
```

FEATURE:  
NAME/KEY: CDS  
LOCATION: (1).....(1848)  
OTHER INFORMATION: HI-0240  
US-10-260-877-119

Alignment Scores:  
Pred. No.: 1,066-35 Length: 1848  
Score: 408.50 Matches: 87  
Percent Similarity: 65.8% Conservatives: 48  
Best Local Similarity: 42.4% Mismatches: 59  
Query Match: 14.70% Indels: 11  
DB: 14 Gaps: 2

US-09-868-987-14 (1-552) x US-10-260-877-119 (1-1848)

```

Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22
Db 1183 AATGTTGCAACAAATTCAAGGACGTTTGGTCTTAATTTCCAAATTTACTGGTGTTCATAGC 1242
Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42
Db 1243 ATTGGGAGACACATAATCTTCTACCTATTGAATCTGGTGCAITTAATTCACCAAT 1302
Qy 43 GluValLeuSerGluGluThrIleSerSerAspLeuLysLysGlnCysThrGlnGly 62
Db 1303 CAATTTGTTGAAGAACGACCAATTTGCCCATCTATTAGTGGCGCAAAACGTAGACGAGG 1362
Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrArg 82
Db 1363 ATTAATGCGAGCTTTTGGGGATTAGTCTGTATTGCTTTATGCTTTTACTACAAA 1422
Qy 83 PheGlyGlyValIleAlaSerGlyValValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 102
Db 1423 ATGTTGTTGTTGATTGCAAGTTTTCACCTTTGTTTATTAATTCGATTACTTTGTGGGATA 1482
Qy 103 LeuGlnTyrLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
Db 1483 ATGCTATTTTACCGCGCGACACTTTCAATGCGGGTATTGCGGGTATCGTTTAACT 1542
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
Db 1543 TTAGTATGTCAGTAGATGCAATGTAITGATTGTTTGAACGTAITTAAGAGAAATTCGT 1602
Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
Db 1603 AATGCTGCTTCAATTCAGACAGCCCAATTAATGAAGGTTTAAACGCGCATTTACTTCTATT 1662
Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
Db 1663 TTTGATGCAAACTTAACCAACATCTTAACCGCAATTAATCTATACGCGGTAGGAACAGGC 1722
Qy 182 ProIleLysGlyPheAlaLeuThrIleLeuLeuGlyIlePheSerSerMetPheThrAla 201
Db 1723 CCAATTCAAGGTTTTCGATTACGCTTTACTTGTGTTGCACTTCTATGTTTACCACG 1782
Qy 202 LeuPheMetThrLys 206
Db 1783 ATTACAGGAACCTCGC 1797

```

# RESULT 9

US-09-881-752A-95  
Sequence 95, Application US/09881752A  
Patent No. US20020115078A1  
GENERAL INFORMATION:  
APPLICANT: Kleantous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the  
TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/041002  
CURRENT APPLICATION NUMBER: US/09/881,752A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 08/833,457  
PRIOR FILING DATE: 1997-04-01  
NUMBER OF SEQ ID NOS: 370  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95  
LENGTH: 1671  
TYPE: DNA  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (50)....(1624)  
US-09-881-752A-95

Alignment Scores:  
Pred. No.: 5,116-34 Length: 1671  
Score: 393.00 Matches: 95  
Percent Similarity: 57.49% Conservatives: 47  
Best Local Similarity: 38.46% Mismatches: 99  
Query Match: 14.15% Indels: 6  
DB: 10 Gaps: 3

US-09-868-987-14 (1-552) x US-09-881-752A-95 (1-1671)

```

Qy 3 SerSerProIleLeuAsnValProLeu---LysAsnHisAlaSerValSerGlyLysPhe 21
Db 932 TCACCCCGGTGATTAGGAGCGTATCGTGGGGGAGCGGAGATTAGCGGAATTTT 991
Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41
Db 992 AGCTGGCTCAAGCAGCGGATTAGCGCATCGCTTTAGGAGTGGGCGCATGAGCGCTCCC 1051
Qy 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThrGln 61
Db 1052 ATTCAGGTTTTAGAAAAAGAAATTTATAGGCCAAAGTTTAGGAAAGACAGCGTTAAAACT 1111
Qy 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 81
Db 1112 TCCATTATCGCTAGTTGGGGCTTTATTTAGTCATGGCTTTATGCTGCTTTATTATAC 1171
Qy 82 ArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuAsnLeuLeuLeuIleTrpAla 101
Db 1172 TCTATGGCGGGGTGATCGCTTTGCGGTAGTGTCAATCTTTTTCATTGTGCGC 1231
Qy 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
Db 1232 GTCATGCGGATTTTGGAGCGACGCTGACTTTACCGGAATGGCGGATTTTAAACC 1291
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
Db 1292 GTGGGATTCGCGGTGATGCTAATATCATCAACGAGCGCATTAGAGAGTCTTAAGA 1351
Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
Db 1352 GAGATGAGGCGATCGCTAAAGCGCATCCATTAGGCTATATCAATGCGAGCGCGCGGAT 1411
Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
Db 1412 TTTGATTTCTAATATCACTTCTTCATCGCTTCACTGTTTATTATACGCTTATGGCACAGA 1471
Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201
Db 1472 GCGATTAAAGGCTTTGCGCTAACTACAGGATTTAGCGATTTTAGCCTCTATTATCACCGCT 1531
Qy 202 LeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisGlnLeu 221
Db 1532 ATTCTTGGCAGCAAGGATTTTCAAGCGCTTTTACTCAAACTC-----ACTCAACA 1585
Qy 222 HisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeu 241
Db 1586 AAAAGCCTTTTACTTTTGGTTGGGTGATGATAA-----AGAGCTTAGGAGGTTT 1636

```

Qy 242 TrpAlaValSerGlySerVal 248  
 Db 1637 TGGAAATTATTCACACGAACTA 1657

RESULT 10  
 US-09-882-227-401  
 ; Sequence 401, Application US/09882227  
 ; Publication No. US20030158396A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kleanthous, Harold  
 ; APPLICANT: Al-Garawi, Amal  
 ; APPLICANT: Miller, Charles  
 ; APPLICANT: Tomb, Jean-Francois  
 ; APPLICANT: Oomen, Raymond P.  
 ; TITLE OF INVENTION: Identification of Polynucleotides  
 ; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the  
 ; TITLE OF INVENTION: Genome  
 ; FILE REFERENCE: 06132/047002  
 ; CURRENT APPLICATION NUMBER: US/09/882,227  
 ; CURRENT FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: US 08/902,615  
 ; PRIOR FILING DATE: 1997-07-29  
 ; NUMBER OF SEQ ID NOS: 638  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 401  
 ; LENGTH: 1035  
 ; TYPE: DNA  
 ; ORGANISM: Helicobacter pylori  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (25)...(993)  
 US-09-882-227-401

Alignment Scores:  
 Pred. No.: 1.74e-29 Length: 1035  
 Score: 350.00 Matches: 91  
 Percent Similarity: 50.81% Conservative: 66  
 Best Local Similarity: 29.45% Mismatches: 104  
 Query Match: 12.60% Indels: 48  
 DB: 12 Gaps: 7

US-09-868-987-14 (1-552) x US-09-882-227-401 (1-1035)

Qy 244 ValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSer 263  
 Db 88 GTTTCAGCAATTTCAGCGTCTAGCGTTGGGGCTTTGTTTTTCAAAGGGTTTCT--- 144  
 Qy 264 ValLeuGlyMetAspPheLysGlyGlyTyrrAlaPheThrPheAsnProLysGluHisGly 283  
 Db 145 ---TTAGGGATTGATTTCGGGGGGG----- 168  
 Qy 284 lleserAspValAlaGlnMetArgGly-----LysValValHisLys 297  
 Db 169 -----AGTTGGTGCAGTGCCTACACTCAAAACGCCGCCCATTAAGAAGTGCAGAT 222  
 Qy 298 LeuGlnGluAlaGlyLeuSerSerArgAspPheArgileGlnThrPheGlySerSerGlu 317  
 Db 223 CTGTTTGAAGAAGAGCTCGCTTCAAAGCGCTGCAAGTGCAGCGAATTTGGCTCTAAAGAA 282  
 Qy 318 LysileLysileTyrrPheSerAspLysAlaLeuSerTyrrThrLysGlnileArgAlaSer 337  
 Db 283 GAAATTTTATCAATTCCTTTTGTAGAACCGCTGAAATGAAGATCTGAACGCT--- 339  
 Qy 338 LeuLeuLysLeuThrilleMetSerTrpArgTyrrCysGlyileValValArgAsnArgPro 357  
 Db 340 -----ATCGTGGGCAAC----- 351  
 Qy 358 ArgPheLeuTyrrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLys 377  
 Db 352 ---ATTCTAAACCCAGCGGGGATTTTGAATCCGTAATTTGACACCGGGGCCCTAGA 408  
 Qy 378 LeuSerLysLysMetArgTyrrGlnAlaThrilleGlyLeuLeuGlyAlaLeuAlailelle 397

Db 409 GTGGGAGCGCAATTGAAAGAGAAAGCAATTTTCGCTGATTTTAGCATTAATAGCGATC 468  
 Qy 398 LeuLeuTyrrValSerLeuArgPheGluTrpGlnTyrrAlaPheSerAlaValCysAlaLeu 417  
 Db 469 ATGGTTTATGTGAGTTTCGCTATGAATGCGGTTTTCGCTTACGAGCGTCATTGCGCTT 528  
 Qy 418 lIleHisAsp-----LeuLeuAlaThrCysAlaValLeuPheileAlaHisPheLeu 435  
 Db 529 GTGCATGATGTGATTTTAGTGGCAAGCTCGGTGATTTT----- 570  
 Qy 436 LysLysileGlnileAspLeuGlnAlaileGlyAlaLeuMetThrValLeuGlyTyrrSer 455  
 Db 571 ---AAGATTGATATGAATTTGGAAGTGAATTCGCGCTTGTCTACCTTGATTGGGTATTC 627  
 Qy 456 LeuAsnAsnThrleullellePheAspArgileArgGluAspArgGlnAlaAsnLeuPhe 475  
 Db 628 ATTAATGATCATCATTAATTTTACAGGATCAGAGAGAGATGCTTCTCAAAAAACC 687  
 Qy 476 ThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMet 495  
 Db 688 AAAACGCCCACTCAAGCCATTGATGAAGCCATTCTAGCAGCTCAGCGCAGCTTTTA 747  
 Qy 496 ThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheileGlyGlySerSerVal 515  
 Db 748 ACTTCTTTAACCGTGTGTTTGTGGGTGTTGATTTCGTCGTTGGAGTAAGATCATC 807  
 Qy 516 PheAsnPheAlaPheileMetThrilleGlyileLeuLeuGlyThrLeuSerSerLeuTyrr 535  
 Db 808 ATTGGCTTTTCATTGCCCATTGTTAATAGCAGCAATTGTAGGGACTTATAGCTCTATTTTC 867  
 Qy 536 lIleAlaProProLeuLeuLeuPheMet 544  
 Db 868 ATCGCCCTAAAGTGGCTTATTGTTTA 894

RESULT 11  
 US-10-260-877-129  
 ; Sequence 129, Application US/10260877  
 ; Publication No. US20030021813A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abbott Laboratories  
 ; APPLICANT: Chovan, Linda E.  
 ; APPLICANT: Hessler, Paul E.  
 ; APPLICANT: Reich, Karl A.  
 ; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME  
 ; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: 'ESSENTIAL GENES'  
 ; FILE REFERENCE: 6565.US.P1  
 ; CURRENT APPLICATION NUMBER: US/10/260,877  
 ; CURRENT FILING DATE: 2002-09-30  
 ; PRIOR APPLICATION NUMBER: US/09/649,145  
 ; PRIOR FILING DATE: 2000-08-25  
 ; NUMBER OF SEQ ID NOS: 137  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 129  
 ; LENGTH: 975  
 ; TYPE: DNA  
 ; ORGANISM: H. influenzae  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(975)  
 ; OTHER INFORMATION: HI-0239  
 US-10-260-877-129

Alignment Scores:  
 Pred. No.: 1.27e-28 Length: 975  
 Score: 342.00 Matches: 98  
 Percent Similarity: 51.90% Conservative: 80  
 Best Local Similarity: 28.57% Mismatches: 111  
 Query Match: 12.31% Indels: 54  
 DB: 14 Gaps: 15

US-09-868-987-14 (1-552) x US-10-260-877-129 (1-975)



Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrCln 220  
 Db 1195 CGTGTCTATGGTTCAGGGTGTGCTAGGCTCTGGGTAACAGC----- 1239  
 Qy 221 LeuHisMetMetAsnLys-----PheValGlyIleLys 231  
 Db 1240 ---AAATTTTAAATAAACGGCTGGCTGGTGTGGCGTGAAA 1278

## RESULT 13

US-08-781-986A-671  
 ; Sequence 671, Application US/08781986A  
 ; Publication No. US20030054436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,445  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 671:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 609 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-781-986A-671

## Alignment Scores:

Pred. No.:	1.69e-25	Length:	609
Score:	311.50	Matches:	62
Percent Similarity:	64.57%	Conservative:	51
Best Local Similarity:	35.43%	Mismatches:	53
Query Match:	11.21%	Indels:	9
DB:	8	Gaps:	3

US-09-868-987-14 (1-552) x US-08-781-986A-671 (1-609)

Qy 378 LeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIlelle 397  
 Db 8 ATTCGACAGGAATAGCAAAAATGCAATGCTTCATTAATCTATGATGATGATGATC 67  
 Qy 398 LeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeu 417  
 Db 68 ATCATCTATGATCATTCATGACGATTTGAATGGCGCATGGTCTTTCATCTGATTGGCATT 127  
 Qy 418 IleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLys 437  
 Db 128 TTATCATGATGATTC-----ATTATAGTAGCGATTTTTCAGTTTATTATTAGA 172

## Alignment Scores:

Pred. No.:	2.41e-22	Length:	595
Score:	293.50	Matches:	62
Percent Similarity:	56.90%	Conservative:	37
Best Local Similarity:	35.63%	Mismatches:	56
Query Match:	10.21%	Indels:	19
DB:	8	Gaps:	2

Qy 438 IleGlnIleAspLeuGlnAlaIlelleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsn 457  
 Db 173 ATTGAAGTAGATTAAACATTTATCGCGCTGTATTACAAATGTCGGTTATTCAATTAAT 232  
 Qy 458 AsnThrLeuIlelePheAspArgIleArgGluAspArgGln---AlaAsnLeuPheThr 476  
 Db 233 GATACATCGTAACGTTTACCGGTGACGTGAAACCTTCAAAAGGTTAAAGTGATTACG 292  
 Qy 477 -----ProMethisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThr 493  
 Db 293 ACAACAGAACAAATTCATGATATATCGTTAATAGATCAATTAGACAGACAATGACACGTTCA 352  
 Qy 494 ValMetThrThrAlaThrLeuSerValLeuMetLeuLeuPheIleGlyGlySer 513  
 Db 353 ATTAATACAGTATTACAGTTATTGTAGTAGTGTATATCTATCTCTTCGGTGTCTCT 412  
 Qy 514 SerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSer 533  
 Db 413 ACGATATTCAACTTTACTTTTAGCATTTATTTATCGGATTGATTTCTCGGTGATTCTCTTCA 472  
 Qy 534 LeuTyrIleAlaProPheLeuLeuPheMetValArgLysGlu 548  
 Db 473 ATCTTCATTGCGCTTCGCTATGGGAATAAATGAAAAACGTCAG 517

## RESULT 14

US-08-781-986A-1128  
 ; Sequence 1128, Application US/08781986A  
 ; Publication No. US20030054436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,446  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 1128:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 595 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-781-986A-1128

## Alignment Scores:

Pred. No.:	2.41e-22	Length:	595
Score:	293.50	Matches:	62
Percent Similarity:	56.90%	Conservative:	37
Best Local Similarity:	35.63%	Mismatches:	56
Query Match:	10.21%	Indels:	19
DB:	8	Gaps:	2

```

US-09-868-987-14 (1-552) x US-08-781-986A-1128 (1-595)
Qy      91 AlavalLeuAsnLeuLeuGlnTyrAlaAlaLeuGlnTyrLeuAspAlaProLeu 110
Db      GCATTGCACACTTATCTATCTAAAGCGTAGTCATTAAATTTCAATTCGGGGGTCTTA 67
Qy     111 ThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaSerVal 130
Db      ACTTTACCAGCATAGCCGGTGTTAGGTATTAGGTGTTAGGTGTTAGGTGTTAGGTGA 127
Qy     131 LeuValPheGluArgIleArgGlnGluPheLeuLeuSerGlnSerLeuLysSerVal 150
Db      ATCATGATGAGCGGATTAAGGATGAACAATTCCGAATAAGTAGAACGATAAGAAGCCTTT 187
Qy     151 GluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThrThrValLeu 170
Db      TCTAAAGCAAACAAAAGTTTCATTCTTAACAAATTTTTGTATCTTAACAAACAGTTATC 247
Qy     171 AlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu 190
Db      GCCGACAGCATATTAATCTCTTCCGTTGAAAGTTCCAGTTAAAGGTTTCGCCACAATGTTA 307
Qy     191 IleLeuGlyIlePheSerMetPheThrAlaLeuPheMetThrLysPhePheMet 210
Db      TTATTAGGTATTCTAATGATCTTTGTTACAGCGCGTCTCTTATCAAGATTCTTATTATCA 367
Qy     211 Leu-----TipMetAsnLysThrGlnHisThr 219
Db      TTACTGTTTCATCAAAATATATTCAAAAAATCAATTTGGTTATTNGGTTAAAAAGAAAT 427
Qy     220 GlnLeuHisMetMetAsnLysPheValGlyIle-----Lys 231
Db      AAACGACATGATATTAATGAAGGTGATAGTTTTCATGACCTTAAAACTTCATTCGAGAAA 487
Qy     232 HisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSer 245
Db      TGGAATTTTGTAAATTAAGTAAAGCCATTAAITGGAGTAAGT 529

RESULT 15
; Sequence 6816, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6816
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1104)
US-10-156-761-6816

Alignment Scores:          Pred. No.:       5,29e-21         Length:        1104
Score:                     Matches:           275,50
Percent Similarity:       Conservative:            86

```

Search completed: October 28, 2003, 18:57:46  
Job time : 12261 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 3261 Seconds  
(without alignments)  
4114.097 Million cell updates/sec

Title: US-09-868-987-14  
Perfect score: 2778  
Sequence: 1 MVSSPILNVLKHNHASVSGK.....SLYIAPLLFLFWVRKENSRSK 552

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgn2\_1/USFTG\_spo3/US09868987/runat\_28102003\_152602\_23844/app\_query.fasta\_1.711  
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09868987 @CNC 1.1.2810 runat\_28102003\_152602\_23844 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_fod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_ges1.\*

29: gb\_gse2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	404	14.5	628	28	AY080093
C 2	388.5	14.0	934	29	AY080093
C 3	275.5	9.9	1189	29	AY080093
C 4	244	8.8	654	14	CB483547
C 5	230.5	8.3	731	28	CB483547
C 6	221	8.0	627	29	CB483547
C 7	213	7.7	661	29	CB483547
C 8	203.5	7.3	602	28	CB483547
C 9	203.5	7.3	602	29	CB483547
C 10	191.5	6.9	1124	29	CB483547
C 11	182	6.6	1348	29	CB483547
C 12	168	6.0	548	28	CB483547
C 13	158.5	5.7	662	14	CB483547
C 14	151.5	5.5	1101	29	CB483547
C 15	150.5	5.4	655	28	CB483547
C 16	150	5.4	1514	29	CB483547
C 17	149	5.4	1305	29	CB483547
C 18	135.5	4.9	1587	29	CB483547
C 19	126	4.5	329	9	CB483547
C 20	122	4.4	297	9	CB483547
C 21	122	4.4	559	14	CB483547
C 22	120	4.3	955	14	CB483547
C 23	120	4.3	924	13	CB483547
C 24	120	4.3	943	14	CB483547
C 25	120	4.3	956	13	CB483547
C 26	120	4.3	1418	29	CB483547
C 27	120	4.3	3431	11	CB483547
C 28	119	4.3	187	28	CB483547
C 29	114.5	4.1	537	14	CB483547
C 30	113.5	4.1	898	29	CB483547
C 31	113	4.1	707	14	CB483547
C 32	113	4.1	3459	11	CB483547
C 33	112	4.0	980	14	CB483547
C 34	111	4.0	1813	12	CB483547
C 35	110.5	4.0	1201	13	CB483547
C 36	109.5	3.9	672	29	CB483547
C 37	109.5	3.9	750	28	CB483547
C 38	109	3.9	818	14	CB483547
C 39	108.5	3.9	2486	29	CB483547
C 40	107.5	3.9	2080	11	CB483547
C 41	107	3.9	676	28	CB483547
C 42	106	3.8	1031	12	CB483547
C 43	105.5	3.8	1331	10	CB483547
C 44	105.5	3.8	1662	10	CB483547
C 45	105	3.8	899	13	CB483547

ALIGNMENTS

RESULT 1  
AY080093/c  
LOCUS AY080093  
DEFINITION AY080093 Scripps Pier (La Jolla, CA) uncultured virus community  
sequence.  
ACCESSION AY080093  
VERSION AY080093.1 GI:24745274  
KEYWORDS GSS.  
SOURCE uncultured marine virus  
ORGANISM uncultured marine virus  
REFERENCE 1 (bases 1 to 628)  
AY080093 628 bp DNA linear GSS 06-NOV-2002  
uncultured marine virus genomic clone SI051P3C7L, genomic survey



```
Db 756 GGC---GGTTTCATCGCCAGGCGCTGAACCTCGGTATCGACTTCCAGGGCGGCTCCGTC 700
Qy 275 PhetrPheAnProLysGluHISglYleSerAspValAlaGlnMetArgGlyLysVal 294
Db 699 ATCGAAGTCCGCGCAACGAG---GGTGAGCGCATCTGGCGATATCCGCGATCGCTA 643
Qy 295 ValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheAsglleGlnThrPheGly 314
Db 642 AACGAG-----CTAAACCTGGTGAAATCCAGGCTCAGAATTTCGGT 601
Qy 315 SerSerGluLysIleLysIleTyrPhe-----SerAspLysAlaLeu 328
Db 600 ACGCGCAGGATGTTCTTATCCGATTTCAGGCGAGGAGGCGGCGAAATCCGAGCAA 541
Qy 329 SerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyr 348
Db 540 TCCGCCATCACCATTGTCGCGCGGAGCTT----- 511
Qy 349 CysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAla 368
Db 510 -----GAGGACAAATATGACTTC 493
Qy 369 LysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIle 388
Db 492 CGCGCGTGAAGTTGTGCGCCCGCCGCTTCCGGTGACCTCACCTTCCACCTCCAGCATC 433
Qy 389 GlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuLeuArgPheGluTrpGln 408
Db 432 GGCATCTTCGTCGCGCATACGCGGCATCATGATCTACATCTGGGTTCGCTTCGAATGCGAG 373
Qy 409 TyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeu 428
Db 372 TTTCGGCTCGCGCGCTCATTTTCGATGTGCAGACGCTGCTTCACA----- 325
Qy 429 PheIleAlaHisPhePheLeuLysLysIleGlnLeuAspLeuGlnAlaIleGlyAlaLeu 448
Db 324 --ATCGCCCTCTTCGTTCTCTGGGATAGAGTTCAACCTTCCAGCATGTCGCGCATC 268
Qy 449 MetThrValLeuGlyTyrSerLeuAsnThrLeuIleIlePheAspArgIleArgGlu 468
Db 267 CTGACGATTATCGGTATTCGCTGAATGACACCGCTGCATCTATGACCGTATCCGGAA 208
Qy 469 AspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLys 488
Db 207 AACCTCAGCGCTATAGAAAGATCGCTGCGATGATATTGAGTGTCTCTGAACGAG 148
Qy 489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508
Db 147 ACGTGTGCGCACCACCATCTCACCAGCTTGACCGTGTCTCTGGCGCTCTCGCTCTAC 88
Qy 509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528
Db 87 CTCCTCGTGGCGAGGTCACTCGCTCTACCTTCGCCCATGCTCTTCGGTGTGCGGATC 28
Qy 529 GlyThrLeuSerSerLeuTyrIle 536
Db 27 GGTGTGTTCTCGTCTCTACATT 4
```

```
RESULT 3
BZ560638 1189 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_2652.xl pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_2652, Genomic survey sequence.
ACCESSION BZ560638
VERSION BZ560638.1 GI:27179032
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1189)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
```

Burns,J.L., Kaul,R. and Olsen,M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 20622216954  
Fax: 2066885244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers

FEATURES  
source

1..1189  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone\_lib="pacs2-164\_2652"  
/notes="clinical isolate 2-164 Whole genomic shotgun  
library."  
BASE COUNT 220 a 361 c 362 g 244 t 2 others

Alignment Scores:  
Pred. No.: 1,27e-19 Length: 1189  
Score: 275.50 Matches: 54  
Percent Similarity: 64.19% Conservative: 41  
Best Local Similarity: 36.49% Mismatches: 40  
Query Match: 9.92% Indels: 13  
DB: 29 Gaps: 4

US-09-868-987-14 (1-552) x BZ560638 (1-1189)

```
Qy 401 ValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
Db 43 GTCGGCTTCGCTTCCAGTTCGAAAGTTCCGCTCGGTCGATCCTCTCGTGGTGCACGAT 102
Qy 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIle 440
Db 103 GCGATCATCGTGTGCGGCTGCTG-----TCGTTCTTC-----CAGGTGACCTTC 147
Qy 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThrLeu 460
Db 148 GACCTGACCTCTCTCGCGCTGTGCTGGCGGTGCTTACTCGCTGAACGACCATC 207
Qy 461 IleIlePheAspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThr 476
Db 208 GTGATCTTCGACCGGTGCGGAGAACTTCGCGTGTGCGCAAGCGCGATCTGGTCGAG 267
Qy 477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496
Db 268 AACCTGAAATC-----TCCACGACGACAGACCTGTGTCTACTATCGCTATCGCCACC 315
Qy 497 ThrAlaThrThrLeuSerValLeuMetLeuLeuPheIleGlyGlySerSerValPhe 516
Db 316 TCGGTTTCCACGCTGCTGCTATCGCGCGCTGCTGTCTTCGCGCGCGACACCTGTC 375
Qy 517 AsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIle 536
Db 376 GGTTTCTCCATCCCTGTTCGTCGGTGTAAATGGCGGTACTTACTCTGTCGATCTACATC 435
Qy 537 AlaProProLeuLeuLeuPheMet 544
Db 436 GCCAACGTGTTGATTCGGCTG 459
```

RESULT 4

CB483547/c  
LOCUS CB483547  
DEFINITION cclwbtb1008006 Coregonus clupeaformis head Coregonus clupeaformis  
CDNA, mRNA sequence.  
ACCESSION CB483547  
EST 01-APR-2003



```

DB:                28                Gaps:                9
US-09-868-987-14 (1-552) x BH397261 (1-731)

QY 114 GlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPhe 133
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 GGTATCGGGGGTATCGTATTATTCATGCCAAGGAGTAGATACGAACGTAATCATTTAT 668

QY 134 GluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLeuLysSerValGluLysGly 153
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 667 GAAAGACCAAGAGAGATTATTCGAGGAAAAGAAATCCGTAAGCAGCATAACACCGAGT 608

QY 154 TyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr-ThrValLeuAlaSerAl 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 607 TTCAAGCAGCGTTTATCTCGGATTATTGACGCACACGCCACACATTACTACCAGGTG 548

QY 173 aleuLeuPhePheLeuAspThrGlyProIleLysGlyPheAla-LeuThrLeuIleLeuG 193
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 547 TTGTATTATATTTCGGTACAGGACCAATCCAAGGATTTCGCGAAACCTTAATTATCG 488

QY 193 lYilePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMetLeuTrpM 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 487 GTATTTTGATGACGTTCTTACTTCTGTATTATTGTCGAGAGTAATGATCTTCACGAC 428

QY 213 etAsnLysThrGlnHisThrGlnLeu-----HisMetMetAsnLysPheValGlyI 230
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 TTGGAAGAGGTAAGAGATTTTCAGTATGGACATCTTTCTCAAAAAATCTTTTCAGAAATA 368

QY 230 leLysHisAspPheLeuArgGlyCysLysLysLeuTrpAla---ValSerGlySerValP 249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 TCTGGATCGACTTTATCGGG-----AAAAGAAAATGGTCTTACATCTTCTACTATAT 314

QY 249 heLeuLeuGlyCysValAla-----LeuGlyPheGlyAlaIlePheMetLeuVal 265
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 313 TGATGATTATCTGTATCGCATCTATTGTTACAAAAGGCTC-----AAAT 269

QY 265 euGlyMetAspPheLysGlyGly-----TyrAlaPheThrPheAsnProLysGluHisG 283
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 TTGGGGTTGACITCAAGAGGAGGAAGTTATGTTGTAAGATTTCAT---AAGCCGGTTG 212

QY 283 lYileSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyL 303
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 TGGCTTCCGATATTCAGGAAGAGTTGGCTCCGATAATA---AACTAAT-GACGCTAAAA 156

QY 303 euSerSerArgAspPheArgIleGlnThrPheGlySerSerCluLysIleLysIle 321
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 ACGAAGCTGTAGT-----GTTAAACATTTGGTAACTCTTAATCAGCTTAGAATA 106

RESULT 6
BZ564006          627 bp      DNA      linear      GSS 17-DEC-2002
LOCUS             pacs2-164_4490.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION        BZ564006
ACCESSION         BZ564006
VERSION           BZ564006.1 GI:27188954
KEYWORDS          GSS.
SOURCE            Pseudomonas aeruginosa
ORGANISM          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
                  Pseudomonadaceae; Pseudomonas.
REFERENCE         1 (bases 1 to 627)
AUTHORS           Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
                  Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE            Whole-Genome-Sequence variation among multiple isolates of
                  Pseudomonas aeruginosa library
JOURNAL           J. Bacteriol., (2002) In press
COMMENT          Contact: Chris K. Raymond
                  Genome Center
                  University of Washington
                  Box 352145, Seattle, WA 98105-2145, USA
                  Tel: 2062216954
                  Fax: 2066857244
                  Email: craymond@u.washington.edu

```

```

FEATURES
    source          location/Qualifiers
    1..627          /organism="Pseudomonas aeruginosa"
                   /mol_type="genomic DNA"
                   /strain="2-164"
                   /db_xref="taxon:287"
                   /clone="pacs2-164_4490"
                   /clone_lib="pacs2-164"
                   /note="clinical isolate 2-164 whole genomic shotgun
                           library."
BASE COUNT        100 a      196 c      182 g      149 t
ORIGIN
Alignment Scores:
Pred. No.:        6.3e-14      Length:      627
Score:            221.00      Matches:    44
Percent Similarity: 65.57%     Conservative: 36
Best Local Similarity: 36.07%  Mismatches:   32
Query Match:      7.96%       Indels:     10
DB:               29          Gaps:         3
US-09-868-987-14 (1-552) x BZ564006 (1-627)
QY 427 ValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGly 446
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 GTGATGGGGGTGCTGCTCTCTTC-----CAGGTGACCTTCGACCTGACCGTCTCGCC 162

QY 447 AlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIle 466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 GCTGTGCTGGCGGTGCTGCTTACTCGGTGAACACACATCGTGTCTTCGACCGGGTG 222

QY 467 ArgGluAsp-----ArgGlnAlaAsnLeuPheThrPrometHisValLeuVal 482
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 CGCGAGAACTTCGCGGTGCTCGCAAGCCGATCTGTGCGAAGCAATC----- 276

QY 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 277 -----TCCACACCCAGACCTGTTGGTACTATCGCACCTCGGTTTCACGGTGTG 330

QY 503 ValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMet 522
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 GCTATCGCGCGGTGCTGTTCTTCGGGGGCAACACCTGTTGGTTCCTCATCGCCCTG 390

QY 523 ThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeuLeuLeu 542
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 391 TTCGTGCGTGAATTCGGGGGTACCTACTCTGTCGATCTACATCGCAACGCTGTTTATG 450

QY 543 PheMet 544
    |||:|||||:
Db 451 TGGCTG 456

RESULT 7
BZ564007          661 bp      DNA      linear      GSS 17-DEC-2002
LOCUS             pacs2-164_4490.x3 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION        BZ564007
ACCESSION         BZ564007
VERSION           BZ564007.1 GI:27188956
KEYWORDS          GSS.
SOURCE            Pseudomonas aeruginosa
ORGANISM          Pseudomonas aeruginosa
                  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
                  Pseudomonadaceae; Pseudomonas.
REFERENCE         1 (bases 1 to 661)
AUTHORS           Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
                  Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE            Whole-Genome-Sequence variation among multiple isolates of
                  Pseudomonas aeruginosa library
JOURNAL           J. Bacteriol., (2002) In press
COMMENT          Contact: Chris K. Raymond
                  Genome Center
                  University of Washington

```









QY 460 uillePheAspArgIleArgGluAsp-----ArgGlnAlaAenLeuPheT 476  
 Db 561 -CTGATCTCGACCGGTCGCAAGACTTTTCGGCGTCTGCGCAAGCCGATCTGGTCG 503  
 QY 476 hrPrometHisValLeuValAenAspAlaLeuGlnLysThrPheSerArgThrValMetT 496  
 Db 502 AGAAGCTTGAACATC-----TCCACGAGCGACACCCCTGTGGTACCATCGCCA 455  
 QY 496 hrThrAlaThrThrLeuSerValLeuMetLeuLeuPheIleGlyGlySerSerValP 516  
 Db 454 CTCCTGGTTCCACGCTGCTGCTATTCGGCGCGCTGCTGTCTTCGCGCGCGCAATCTGT 395  
 QY 516 heAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyri 536  
 Db 394 TCGGTTTCTCCATCCGCCCTGTTCGTCGGTGTATGCGCGGTACTACTCTGTCGATCTACA 335  
 QY 536 leAlaProPheLeuLeuPheMet 544  
 Db 334 TCGCCAACGCTGATTGATCTGGCTG 309

RESULT 12  
 AZ303132  
 LOCUS  
 DEFINITION  
 GSSBr2079 Brucella abortus random genomic library Brucella  
 melitensis biovar Abortus genomic clone UU2079, genomic survey  
 sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 GSS.  
 AZ303132.1 GI:10129343  
 Brucella melitensis biovar Abortus (Brucella abortus)  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bruciellaceae; Brucella.

REFERENCE  
 AUTHORS  
 Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,  
 Faccio,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasch,A.C.C.,  
 Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.  
 Gene discovery through genomic sequencing of Brucella abortus  
 Infect. Immun. 69 (2), 865-868 (2001)

JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 21101034  
 11159979  
 Contact: Siv Andersson  
 Small Genomes Sequencing Group  
 Department of Molecular Evolution, Uppsala University  
 Norbyvagen 18C, S-752 36, Uppsala, Sweden  
 Tel: 46-18-471-4379  
 Fax: 46-18-471-6404

Email: Siv.Andersson@ebc.uu.se  
 Sequences were basecalled with phred and vector was masked with  
 crossmatch (see <http://genome.washington.edu>). Sequences were then  
 trimmed from both ends to remove low quality bases and masked  
 vector.

Clas: shotgun.

FEATURES  
 Location/Qualifiers  
 1..548  
 /organism="Brucella melitensis biovar Abortus"  
 /mol\_type="genomic DNA"  
 /strain="2308"  
 /db\_xref="taxon:235"  
 /clone="UU2079"  
 /clone\_lib="Brucella abortus random genomic library"  
 /note="Vector: modified M13"  
 BASE COUNT 98 a 158 c 156 g 136 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.32e-08 Length: 548  
 Score: 168.00 Matches: 32  
 Percent Similarity: 66.29% Conservative: 27  
 Best Local Similarity: 35.96% Mismatches: 30  
 Query Match: 6.05% Indels: 0  
 DB: 28 Gaps: 0

US-09-868-987-14 (1-552) x AZ303132 (1-548)  
 QY 454 TyrSerLeuAenAsnThrLeuIlePheAspArgIleArgGluAspArgGlnAlaAen 473  
 Db 2 TATTCGCTGAACGATCGGTGATCTATACCGGTGCGGAGAACTTCGCGGTAC 61  
 QY 474 LeuPheThrProMetHisValLeuValAenAspAlaLeuGlnLysThrPheSerArgThr 493  
 Db 62 AAGAGTCGCCATTCGCCGCGCATCATCGATCCCTCGAGGGCCAGACCTGTCCGGAACG 121  
 QY 494 ValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySer 513  
 Db 122 CTGTTGACCTCTTTCGTCACCTTCCTGCGACATGTTCCGCTTTATGCTTTTGGCGGCTCG 181  
 QY 514 SerValPheAenPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSer 533  
 Db 182 GAAATTCGCATGTTTGGCGTCGCGTCAGTCTCGGTATCATTTGTGGCGAGTTATTCTTCA 241  
 QY 534 LeuTyriLeAlaProProLeuLeuLeu 542  
 Db 242 ATCTTCATCGCAGCACCGCTGCTTGT 268

RESULT 13  
 CB852238  
 LOCUS  
 DEFINITION  
 UI-CF-FN0-aff-i-17-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone  
 UI-CF-FN0-aff-i-17-0-UI 3', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 CB852238.1 GI:30047205  
 EST.  
 Homo sapiens (human)  
 Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 1 (Bases 1 to 662)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 8889548

Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems  
 ([www.openbiosystems.com](http://www.openbiosystems.com)).  
 Seq primer: M13 FORWARD  
 POLYA=No.

FEATURES  
 Location/Qualifiers  
 1..662  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-FN0-aff-i-17-0-UI"  
 /tissue\_type="Human Lung Epithelial cells"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-FN0"  
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-FN0 is a subtracted cDNA library derived from two  
 normalized Human lung epithelial cell libraries (EN1 and  
 DU1) The library was subtracted according to according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806.

1996. For additional information, contact:

bento-soares@uiowa.edu

TAG SEQ=None found"

BASE COUNT 101 a 215 c 227 g 117 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 6,47e-07 Length: 662  
Score: 158.50 Matches: 57  
Percent Similarity: 41.04% Conservative: 46  
Best Local Similarity: 22.71% Mismatches: 86  
Query Match: 5.71% Indels: 62  
DB: 14 Gaps: 9

US-09-868-987-14 (1-552) x CB852238 (1-662)

Qy 245 SerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerVal 264  
Db 1 CGCGCGCGCGGTTCCTCGTCACCGCGTGCCTGCATTTC-----42  
Qy 265 LeuGlyMetAspPheLeuGlyGlyTyAlaPheThrPheAsnProLysGluHisGlyLe 284  
Db 43 ---TCCATCGAATTCACCGCGCGGACCGCTCATC-----GAGGTGGCCTAC 84  
Qy 285 SerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSer 304  
Db 85 CGCGAGGCGCGAGATCGAG---AAGACCGCGCACACCATCGAGAGATGGGTATGGC 141  
Qy 305 SerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLeuTyPhe--- 323  
Db 142 ---GAGGTCCAGGTGCAGAACTTCGGTCCCTCGCGACGTCATCGCGCTCGCG 195  
Qy 324 ---SerAspLysAlaLeuSerTyThrLysGlnIleArgAlaSerLeuLeuLys 340  
Db 196 GTGGTCCCGCGAGACAGACAGAGTGTGGCAAGGTTCCTGGTGAGCTCTGGCG 255  
Qy 341 LeuThrIleMetSerTrpArgTyCysGlyIleValValArgAsnArgProArgPheLeu 360  
Db 256 GCCGAG-----GCCGCGCGCGTCAGCGAGCACCGAAGTACCGGCC 297  
Qy 361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLys 380  
Db 298 CAGGGC-----GAGCAGGTTCAGCAAGCAGATCTGCAAG 330  
Qy 381 Lys-----381  
Db 331 AAACCCGACGGCGGAGCCCATCGGTGTGGCGCTCCGAGGTGTGGTCCGGCCGTC 390  
Qy 382 ---MetArgTyGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeu 398  
Db 391 GGCGCGGAGTGTGGCCAGGCGCCCAAGCGCTGGCGGTGACGGCGGCGCATCATG 450  
Qy 399 LeuTyTrpValSerLeuArgPheGluTrpGlnTrpAlaPheSerAlaValCysAlaLeu 418  
Db 451 ATCTACCTGGCCATCCCGCTTCGAGTGAAGTTCGCGTGGCGGCATCATCGCAACCTG 510  
Qy 419 HisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysIle 438  
Db 511 CACGACNGTG-----GTCATCATCTGGGCTTCCTCGGCTTCCTCCAGTG 555  
Qy 438 eGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyTrpSerLeuAsnAs 458  
Db 556 GGAGTTCCTGCGGTGTGGCGGCGGTCGCGGTTCGCGGTTCGCGGTTCGCGTCAACGA 615  
Qy 458 nThrLeuIleIlePheAspArgIleArgGlu 468  
Db 616 GTCGGTGTGTCATCTTCAGCCGGTGGCGGAG 646

RESULT 14

CNS008NJ

LOCUS 1101 bp DNA linear GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #

BAC17D15 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL052099

AL052099.1 GI:4933951

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

TITLE

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Location/Qualifiers

1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BAC17D15"  
/clone\_lib="RPCI-98"  
/note="end : T7"

BASE COUNT 183 a 345 c 320 g 199 t 54 others

ORIGIN

Alignment Scores:

Pred. No.: 7,98e-06 Length: 1101  
Score: 151.50 Matches: 37  
Percent Similarity: 50.76% Conservative: 30  
Best Local Similarity: 28.03% Mismatches: 64  
Query Match: 5.45% Indels: 1  
DB: 29 Gaps: 1

US-09-868-987-14 (1-552) x CNS008NJ (1-1101)

Qy 1 MetValSerSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGly 19

Db 616 GTGATCGAGCCCGCGTTCATCCGCGCATTCACCGGCGGCGGCGGCGGCGGCGGCGG 675

Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39

Db 676 GGGTTGAGCGCACGAGCGGCCACCGACATGCGCGCTCTGTCGCGGCGGCGGCGGCGG 735

Qy 40 PheValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCys 59

Db 736 GCACCGCTGACGTCGATCGAGCAGCGTTCATCGCGCGGCGGCGGCGGCGGCGGCGG 795

Qy 60 ThrGlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerVal 79

Db 796 CGCGCGGCGGCGGTACAGCCTTCTGTCGCGGCTTCGTCGTCGTCGTCGTCGTCGTCG 855

Qy 80 TyrTyArgPheGlyGlyValIleAlaSerGlyAlaValIleLeuAsnLeuLeuLeu 99

Db 856 TTCATTGGCGGTTCGCTGGTACGCGAATTCGCTCTGCGCAACCTGGGTCGTCGTCG 915

Qy 100 TrpAlaLeuGlnTyTrpLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 5913 Seconds  
(without alignments)  
3819.062 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKHNHASVSGK.....SLYIAPPLLLFMVRKNSK 552

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPTO\_spool/US09868987/runat\_28102003\_152601\_23835/app\_query.fasta\_1.711  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptlo -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09868987 @Cgn\_1\_1\_3635 @runat\_28102003\_152601\_23835 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2730	98.3	12198	1	AE002179 Chlamydia
2	2730	98.3	14134	1	AE001641 Chlamydia
3	2730	98.3	110000	6	Continuation (7 of
4	2730	98.3	300550	1	AP002547 Chlamydia
5	2088.5	75.2	304769	1	AE016994 Chlamydia
6	1937	69.7	9989	1	AE001318 Chlamydia
7	1924	69.3	12442	1	AE002341 Chlamydia
8	817.5	29.4	10046	1	AE009114 Agrobacte
9	817.5	29.4	10066	1	AE008079 Agrobacte
10	766.5	27.6	349619	1	AP002996 Mesorhizo
11	766	27.6	300600	1	AP005952 Bradyrhiz
12	762.5	27.4	14652	1	AE004799 Pseudomon
13	739.5	26.6	300861	1	AE016777 Pseudomon
14	736	26.5	7687	1	AE012971 Chlorobiu
15	736	26.5	311249	1	AE016860 Pseudomon
16	734.5	26.4	10963	1	AE013703 Yersinia
17	734.5	26.4	205050	1	AJ414155 Yersinia
18	725	26.1	282183	1	CJ11168X4 Campyloba
19	722.5	26.0	3000	1	PCU69979 Rhodobacter
20	719	25.9	13818	1	U32710 Haemophilus
21	719	25.9	110000	6	Continuation (3 of
22	713.5	25.7	11527	1	AE010580 Fusobacte
23	711	25.6	305961	1	AE016937 Bacteroid
24	709.5	25.5	15058	1	AE000652 Helicobac
25	709.5	25.5	19836	1	AE001567 Helicobac
26	707.5	25.5	14261	1	AE004160 Vibrio ch
27	699.5	25.2	300956	1	AE016963 Coxidia
28	699.5	25.2	301442	1	AE016798 Vibrio vu
29	696	25.1	10384	1	AE011297 Leptospir
30	695	25.0	10611	1	AE015748 Shawanell
31	694	25.0	4435	1	ECSECF Escherichia
32	694	25.0	10323	1	AE005220 Escherich
33	694	25.0	10577	1	AE000147 Escherich
34	694	25.0	13544	1	AE015067 Shigella
35	694	25.0	139818	1	U82664 Escherichia
36	694	25.0	292309	1	AE016979 Shigella
37	694	25.0	295741	1	AP002551 Escherich
38	693	24.9	300817	1	AE016756 Escherich
39	692.5	24.9	11540	1	AE009509 Brucella
40	691	24.9	20388	1	AE008714 Salmonell
41	691	24.9	268050	1	AL627266 Salmonell
42	691	24.9	300029	1	AE016842 Salmonell
43	689.5	24.8	11911	1	AE014430 Brucella
44	689	24.8	12201	1	AE006057 Pasteurel
45	687.5	24.7	301235	1	AP005075 Vibrio pa

# ALIGNMENTS

RESULT 1

AE002179 12198 bp DNA linear BCT 30-MAY-2000  
LOCUS Chlamydomophila pneumoniae AR39, section 15 of 94 of the complete  
DEFINITION genome.  
ACCESSION AE002179 AE002161  
VERSION AE002179.2 GI:8163380  
KEYWORDS  
SOURCE  
ORGANISM Chlamydomophila pneumoniae AR39  
Chlamydomophila pneumoniae AR39  
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
REFERENCE 1 (bases 1 to 12198)  
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39  
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
MEDLINE 20150255  
PUBMED 10584935  
REFERENCE 2 (bases 1 to 12198)  
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
COMMENT Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1..12198  
/organism="Chlamydomophila pneumoniae AR39"  
/mol\_type="genomic DNA"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/notes="synonym: Chlamydia pneumoniae AR39"  
145..1521  
/gene="CP0178"  
145..1521  
/gene="CP0178"  
/notes="similar to GB:L26051 SP:P33986 PID:415662;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylglucosamine  
1-carboxyvinyltransferase"  
/protein\_id="AAF38052.1"  
/db\_xref="GI:7189109"  
/translation="MQIAQVGGCGRLNGEVKVGAKNAATKLIVASLLSDQKTLRNV  
PDGDSVLVELCKSLGAVSWDKETEVLEIYPTQCTRVPTFNVNRIPIILIGA  
LLGRCPGVVPTVGGDAIGERTLNPHFGLKQGVIGDSSDGYAKAPRGKGNVI  
HLPPYSGATENILIAIHAKGRTVIKVALEAIEDLVFLQKAGADITDNDRTID  
IFGFGGSGVDHTILPKIEAAGFMAAVSGGVFVFNKOBLLIPFLKMLRSIGGG  
FLVSEGIIEFQRLPVGLVLETDHPGLTDQOPFVLLSQAGSSVTHETVHN  
RLGVHLQWGMGEQCLPHOCLSTKACRYAIGNFPSSAVIHGATPNAWSHLVDPDRA  
GFAYMAALTAEGGSIIEHTLDRGTNNVWGLRSLGAKIQIDFMEQBELTSPKS  
LAURDASL"  
complement(1475..3160)  
/gene="CP0179"  
complement(1475..3160)  
/gene="CP0179"  
/notes="similar to PID:1001350 PID:1001346; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="arginyl-tRNA synthetase"  
/protein\_id="AAF38053.1"  
/db\_xref="GI:7189110"  
/translation="MSTLLSILSVICSQAIAKAPFNLEDWAPETPTKEHFGHYQCN  
DAMKLARVLKAPRAEAIAVELPOEPFSLIAGAGFINFTFSPVFNQQLHEFKD  
ALKLGFQVSPKPKIIIDFSSPNIAKDMHVGHLRSTIIGDSLARIFSVVGHVLRNLHI

GWGTAFGMLITVLOENPCDSDLEDLTSLYKAYVCFTNDEBFKRSQONVVALQAK  
DPOAIWEEKICETSEKAFQKLYDILDIVVERGESFYNPFLPEIILEDKGLLTVS  
NDAKVPHEAFSIPFNWQSGGYNVATDDLAMRYRIEEDHADKILIIIVTDLQCSLHF  
QLEATAIAGALPGIFSHVGLVDPOGKKLKTRSGENKVRKLELLODTAIEKAEA  
LREHRPELTDEAIOERAPVIGAINAKYSDLSHRTSDYFSEKMLRFEGTAMFLLY  
AVVRIQIKRRLIGISQLEGGPEIOEPAEELIALLTLRFPEALESTIKELCPHFLTD  
YLYNLTHKNGFFRDSHIQDSPYAKSRFLCALAEQVLAQVLAQVLAQVLAQVLAQV  
complement(3165..3803)  
/gene="CP0180"  
complement(3165..3803)  
/gene="CP0180"  
/notes="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative; identified by match to PFAM protein family HMM  
PF01553"  
/codon\_start=1  
/transl\_table=11  
/product="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative"  
/protein\_id="AAF73638.1"  
/db\_xref="GI:8163381"  
/translation="MIFRICKFTVWVAFSLFKLYGVYKKNFIKGPALIAVNHNSFL  
DPIALHMCVHECIYHLARSLFNIPMLWKQWCFVVRQDEGNSAFKISRDNKRKK  
LVVIEGARSQDQGPQGVIGIMMAKSRVPIIPVYIRGTFAFNRHOKIPHWKTI  
TCVFGTMYFDDIIQNPEIKNKETYQIITNQ"MNKIAELKAWYESGCKGDP"  
complement(3800..4450)  
/gene="CP0181"  
complement(3800..4450)  
/gene="CP0181"  
/notes="similar to SP:P23863 PID:42839 GB:U00096  
PID:1651431 PID:1651438; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytidylate kinase"  
/protein\_id="AAF38054.1"  
/db\_xref="GI:7189111"  
/translation="MIITIDGPSGTGSTTAKALADHLHFNCTGMYRTIAYARLQ  
SPWATLPLTKLEERPSPFATQPLESPFNHLLTSELTTQEVANAASELSOLPEV  
RAFMQDLQRYAQLGNCVPEGRDMGSKVFPNADLKIFLTSSPEVRAQRRLKDLPEGL  
SPEQIAELVKRDAQADRAHPLVIPENGIVIDSSDLTIQVLEKILALLFRNEL"  
complement(4447..5373)  
/gene="CP0182"  
complement(4447..5373)  
/gene="CP0182"  
/notes="similar to GB:M1330 SP:P06466 PID:1208947  
PID:145476 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidate cytidyltransferase, putative"  
/protein\_id="AAF38055.1"  
/db\_xref="GI:7189112"  
/translation="MLNKNKSKTKGAYGDLQFQVVVHSLVTLFVLLLYSDFPLTS  
FALGTITACGAVGYEYSMAKAMHYPLSTFSAIGSFLFALGSLRWGHSPLPGF  
PDALPWTLLIWMVMSIIFVRKSTIGALQSGVTLFSLVGVIRLPLHVLVSYFTL  
QEPYLIWASLIATTKGADIFGYFGKAFGNKKAQIYSNKTUVGVACGLGATL  
ISFIFLQIPTRFASYPFPALIFELGALGIFGTFGFIIESIFKRAHAKHNSKUKA  
VCGMLDLDLSLLSLPIAYLFLITQSKFEIG"  
complement(5374..6159)  
/gene="CP0183"  
complement(5374..6159)  
/gene="CP0183"  
/notes="similar to PID:1208946 GB:U00096 SP:Q47675  
PID:1552751 PID:1786371; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="undecaprenyl pyrophosphate synthetase"  
/protein\_id="AAF38056.1"  
/db\_xref="GI:7189113"  
/translation="MTYALINDPVDLSLATNNAESKFFSLQRLPNHVAIIMDGNRRWY  
KHRECCGHTTSGHYGAKVNLINAVLDIGIKVLTLYTFTSTFGPKKEIOEIF  
NIFYTQDLQKPLYMENIEICIGDLSKLPKGIQTKINHVSMTASFSRLELVLAVN



```

Qy 400 TyrValSerLeuAtrGpheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419
|||
Db 11664 TATGTAGTTTGGCTTTGCAATGCAATATGCTTTCAGTGCCTATGCGCTTTAATTCAT 11723

Qy 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGln 439
|||
Db 11724 GACCTTTTGGCTACTCGTGTGCAGTCTTGTATTATAGCACATTTCTTTTGAAGAAAATCAA 11783

Qy 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThr 459
|||
Db 11784 ATAGATTTCGAAGCAATTTGGTGCTTTAATGACTGTATTGGGTATTCATTAAACAATACT 11843

Qy 460 LeuIleLeuPheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
|||
Db 11844 TTGATCATTTTGTATGCTATTCGTTGGAAGATCGCCAAGCGAACCTGTTTACCCCTATGCAT 11903

Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
|||
Db 11904 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTACGCCGACGGTAATGACAAACAGCTACA 11963

Qy 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAla 519
|||
Db 11964 ACTCTATCAGTTTGTGTAATGCTTTGTTTATAGCGCGCTCCTCTGCTTTAATTTTGA 12023

Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
|||
Db 12024 TTTTATTATGACCATAGGATTTCTTCTAGGAACCTTTATTCGCTCTTTATATGACCACT 12083

Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
|||
Db 12084 CTGTGTGTTTATGTCGTCGTAAGAAATCGCTCAAAA 12122

RESULT 2
AE001641/c 14134 bp DNA linear BCT 01-DEC-2000
LOCUS AE001641
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.
ACCESSION AE001641 AE001363
VERSION AE001641.1 GI:4376855
KEYWORDS Chlamydia pneumoniae pneumoniae CWL029
SOURCE Chlamydia pneumoniae pneumoniae CWL029
ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 14134)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206506
PUBMED 10192388
REFERENCE 2 (bases 1 to 14134)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. 14134
/organism="Chlamydia pneumoniae pneumoniae CWL029"
/mol_type="genomic DNA"
/strain="CWL029"
/db_xref="taxon:115713"
/complement(158..1924)
/gene="recJ"
/note="synonym: CPn0563"
/complement(158..1924)
/gene="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAD18703.1"
/db_xref="GI:4376856"
/translation="MTNSDNASAGLLWHPKEDPAFLGMIKEFHPPTVAQIFSR
GFQTIQIHKFLYSLSSLYDPGLFLDMSKAVERLLARDKEHVMYIGDSVDGWTG

```

```

VALLVEFLRIDVHVSFFLGAILLKOHGETSTLIAKLEEGITLLITVDCGITAGKEV
SDITROGIDVITTDHMPGKIPHCVCATINEKLDRHTYNRBLTGVGVAARGLVNL
ALISRLNPKSGSLKLLDLVLTGLTITDVGVLNGNRVVRVYGIKEIARGARPGLNK
LCAICVKESEVSTLTVIKIAPKUNSLGRIIDPAKGVLELLLTQDDERDALIMELDN
INNERORIEAEVQDVEILNENPEILKQAAIVLSTAWHARVPIIISARLAKTYNKP
VITAIORIGIGKSARTIGSFFLLGLVKKCSLLISYGGHDFAAVYIMKEDKVEDFKK
QVYKVLNSSLKGGDTLPHLEIDAYADFDAIEYDILLASMELEFPFGKGNLMPYFSKVR
QVYKVLNSSLKGLYLSQKERNLEGVAFGLGRHADALAKASWHYPLEIATYTPRLSQT
GSGVHLVDRFRISSEPRFS"
complement(2010..6218)
/gene="secD/secF"
/note="synonym: CPn0564"
/complement(2010..6218)
/gene="secD/secF"
/codon_start=1
/transl_table=11
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AAD18704.1"
/db_xref="GI:4376857"
/translation="MKQVKNRFAIIICVFALALVYVLPCLYLYAKPDKKIDGNEAE
HIKFTKQQRKQVDIIPRVSAISSLHLFGHIQHPALPIDVSVRFKGGDAEDFI
GNLVGEPNVPKISARLHVGYSRBDDHVIQVASSINTSLVESDFSVYSSENOE
MASSILQRYVSACTTFFKQDCSCYPSIWETAPKEQLQYAKNLSGFFVFSRSLAF
CQOFSNODRLAFILSLSDAAIDVEIQKLLKSVYETLSQACIRSLDCPYIEG
LRDCESSLFFSSIEYCPKPKKIPLTLHSDULLAORTSLSKQRDLDFSRLAKEOKL
SKNLTVOVEDYNGFSFOMMDXDTQKIILOGERLLOGIAELTALTILHRAESCDL
IENPFVFCQPRESEAFGCYIPSNPTDCKHFSKGSVYILLKGLSIVAKYQOQGGKE
LOSFKDLQNLNCFSTEAISWITLGEDQVLEIRHPLOQFLDVGWGGFVIGKEGCAF
EYKDIQDLATVNIQEKNSDLVRWHEQYHAKSMDOERLSAPIYONIFLENMK
LNMRKFSRGENILRLIGIDFVGQRQLLSFKHQKQLTDKIDILKYSDBLCARLNKLG
VSEIELRREGDYIHLSPGSSSTISSEILGTSKMGPHVNVNFKSSVSAKRYEQRFLD
YLMFTSOAQGKTSPEIINTFASALFNEEDVDPSPVHEALTILKLSGLASPSGCEPS
TOLDTFSMIALGKDAEQKANPLVIVFRNALDGLASLDIRPEFAAGEGYVLFNSVD
TSPKNAEKLSPSTFHTWTSAYCOEGISGTANGQYSANRGMVAVIDGYWSSPIL
NYPKNAHSVSGKFTTHREYSLKSLGKSGMSFVPEVLESEETISSDLGKKQGTGIS
ACCGAMLIVLMVYVYFGGVYFAGVAGLNLILLIWAALQYLDAPLTLSLGLAGTILAMG
MAVDANLVFERIEEFLQSLSQSKSVKGYTKAFCAI FDSNLTTLASALLFLDTG
PIKGFALTILIGIESMFTALEWTKFFFLMNMKTQHTOLHNMNKFVGKIKHDFLRGCK
KLWAVSGSVLLGCVAGLGSWNSVFGMDFGGYATTFNPKHGISDVAQMRGKVHVK
LQEAGLSSRDFRIQTFGSEKIKIYFSDKASITKADTSLSPKINDHELAVALGULSE
TGIDFTETNETONFWSKSVSKSKMRYQATIGLGLALAILLYLSURFEMQVAFS
AVCAIHDLLATCAVLFAHFLKFIQIDLQAIQALMTVLGSLNLTLLIIFDRIDR
QANLFTPHVLVNDALQKTFSTVTMTATTLSVLLMLLFISSGVNFNPAFIMTIGILL
GTLSSLIYAPLLLFWVRKESK"
complement(6398..7498)
/gene="CPn0565"
/complement(6398..7498)
/gene="CPn0565"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISISDLIPTQVMVIRWGGNVHYVVPNAQNLPKILGGVL
ACFGALLGCAAFAGVCTIIPCIGLMLGLVLLGFAVLOYSGKMSRERPLFRFTK
VPEKINMLGCLSLLOSNNKIPRCYCYHPCQVEICEGSEIYVTKIPOKSDRNTSI
FLIQEMDQALRGHIEKLSLRKTFADIPSVVSSLSERQEEQQYLDPKVLSWSSSED
QASDRTHPKSAIYVNISDAQEPQRCYDAYTKAFFVTVDQIGDPNIVKHTIYVLT
PILGVDPDALPKEENKLLSQAFLYSAEQVAKRMREKQDSIRIKFTFTDPTSPTS
LYFSPHSGSTPHSVTPISLSGFGVGEQESTPA"
8009..8761
/gene="yaeS"
/note="synonym: CPn0566"
/8009..8761
/gene="yaeS"
/codon_start=1
/transl_table=11
/product="YaeS family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAESKFPQLRQLPNHVAIIMDGNRRYKXKHEECGHTH
TSCHYVGAKVLPNLIINLAVLDLGIKVLITLYTFTENFGPKBEIQIENIYFQLDKQL
PYLMENEICLRICIGDLSKLPKGIQTKINHVSMTASFSLRELVLVAVNYGKGDELVRAF

```





```
QY 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuLeuGlyPheGlyAla 260
Db CTTTGGGCTGTTCTGGAAGTGTCTTTCTTTAGTTGGTCTCTCGGGTTTGGAGCC 2892
QY 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280
Db TGGAAATTCCTGTTTGGGAATGGATTTAAAGAGGGGTATGCTTACCTTTAAATCCAAA 2832
QY 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db GAGCATGGCATCAGGATGTTGCTCAAAATGGTGGCAAAAGTTGTGCATAAACTACAGAA 2772
QY 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320
Db GCTGGTCTTTCTCTAGAGACTTCGGTATTCAAACTTTGGATCTTCAGAAAAAGATCAAA 2712
QY 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340
Db ATCTATTTTAGTGATAAGCTTTAAGCTATATACTAAAGCAGATACGAGCCTCTCTCTAAA 2652
QY 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360
Db ATTACGATCATAGCTGGCGTTAGCTGGGATTTGTTGTCAGAAACAGGCTAGATTTTC 2592
QY 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379
Db TCTACGGAACTCTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAAACTATCG 2532
QY 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399
Db AAGAAATATGCGTTATCAGCGCAACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTCTC 2472
QY 400 TyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419
Db TATGTGAGTTTGGCTTTGAATGCAATATGCTTTTCAGTGGCGGTATGCTTTTAATTCAT 2412
QY 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGln 439
Db GACCTTTTGGCTACCTGTGCAGTCTTGTATTATAGCACATTTCTTTTGAAGAAATTCAA 2352
QY 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThr 459
Db ATAGATTGCAAGCCATTTGGTCTTTAATGACTGTATTGGGGTATTCAATTAACAATACT 2292
QY 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db TTGATCATTTTGTATCGTATTCGTGAAGATCGCCAGCGAACCTTGTTTACCCCTATGCAT 2232
QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db GTTTTAGTTAATGATGCCCTTCAAAAGACGTTTCAGCCGACGGTAATGACAAACAGCTACA 2172
QY 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAla 519
Db ACTCATTCAGTTTGTAAATGCTTTTGTATTATAGCGGGCTCCTGTCTTTAAATTTGCA 2112
QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db TTTATTATGACCATAGGATTTCTTAGAACTTTATCGTCTCTTTATATATGACCACT 2052
QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db CTGTGTGTTTATGTTATGGTCCGTAAGAAATAATCGCTCAAAA 2013
RESULT 3
AR310754_06/c
WPCOMMENT
Sequence split into 13 fragments LOCUS AR310754 Accession AR310754
Fragment Name Begin End
AR310754_00 1 110000
AR310754_01 100001 210000
AR310754_02 200001 310000
AR310754_03 300001 410000
```

```
AR310754_04 400001 510000
AR310754_05 500001 610000
AR310754_06 600001 710000
AR310754_07 700001 810000
AR310754_08 800001 910000
AR310754_09 900001 1010000
AR310754_10 1000001 1110000
AR310754_11 1100001 1210000
AR310754_12 1200001 1230025
Continuation (7 of 13) of AR310754 from base 600001 (AR310754 Sequence 1 from patent US
Alignment Scores:
Pred. No.: 1,11e-237 Length: 110000
Score: 2730.00 Matches: 551
Percent Similarity: 99.28% Conservatives: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 98.27% Indels: 3
DB: Gaps: 0
US-09-868-987-14 (1-552) x AR310754_06 (1-110000)
QY 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db 61906 ATGGTCAGCAGCCCTATTTTAAACGTCCTTCAAAAATCATGCCAGTGTCTCAGGAAA 61847
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 61846 TTTACCCACCGTAGTGAGCAAACTCGCCCTCAGATTTAAAATCTGGAGCGATGTCTTT 61787
QY 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
Db 61786 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACA 61727
QY 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
Db 61726 CAAGGACATATATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTTGTATGAGCGTATAT 61667
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuLeu 100
Db 61666 TATAGATTGGAGCGTATCGCTTCGGGAGCTGTTCTCTGAATCTTTTCTTATCTGG 61607
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 61606 GCAGCTCTACAGTATTTGGATGGCCACTCACCTTGTGAGGACTCGCTGGGATTTGTCTT 61547
QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
Db 61546 GCTATGGGATGGCCGTAGATGCAAAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTT 61487
QY 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 61486 TTAATGTTCAAAAGTCTTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 61427
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 61426 ATTTTGTATTAACCTTGACTCAGATTTGGCCTCAGCACTTCTTTTCTTCTCTAGATACA 61367
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Db 61366 GGGCCTATTAAAGGGTTTCTTTGACATTTGATTTTAGGAATTTTCTTCTCAATGTTTACG 61307
QY 201 AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220
Db 61306 GCTCTTTTTCATGACTAAAATTTTCTTCATGCTGTGGATGNATTAAGACCCCAACATACACAG 61247
QY 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db 61246 TTGCATATGATGAATAAGTTTCGGGGATAAAGCATGATTTCTTGAGAGGATCAAAAAA 61187
QY 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 61186 CTTTGGGCTGTTTCTGGAAAGTGTTTTCTTTTAGGTTTCGTTGCTCTCGGGTTTGGAGCC 61127
```

QY 261 TrpAenSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280  
 Db 61126 TCGAATTCGTTTGGGAATGATTTAAAGAGGAGGTATGCTTTTACCTTTAAATCCAAA 61067  
 QY 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
 Db 61066 GAGCATGCGATCAGCGATGTGCTCAAAATGCGTGCAAAAGTTGTGCATAAACTACAGGAA 61007  
 QY 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
 Db 61006 CTGCGTCTCTCTAGAGACTTCGCTATTCAACATTTGGATCTTCAGAAAAGATCAAA 60947  
 QY 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340  
 Db 60946 ATCTATTTAGTGATAAAGCTTTAGCTATATACTAAAGCAGATACGAGCTCTCTCCTAAA 60887  
 QY 340 sleuThrIleMetSerTyrArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360  
 Db 60886 ATTACGATCATGAGCTGCGGTAGCTGTGGATTTGTTGAGAAACAGGCTTAGATTTC 60827  
 QY 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTyrSerLysValSerSerLysLeuSer 379  
 Db 60826 TCTACGGAACTCTAAACGAAACGCAAAATTTTGTCAAGGTAAAGCAGCAACTATCG 60767  
 QY 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399  
 Db 60766 AAGAAATGCGTTATCAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC 60707  
 QY 400 TyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419  
 Db 60706 TATGTGAGTTTGGCGTTTGAATGGCAATATGCTTTTCAGTGCGGTATGCGCTTTAATTCAT 60647  
 QY 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGln 439  
 Db 60646 GACCTTTGGTACCTGTGCACTGTGTTATAGCACATTTCTTTTCAAGAAATTCGA 60587  
 QY 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLysAsnThr 459  
 Db 60586 ATAGATTTGCAAGCATTGGTCTTTAATGACTGTATTGGGTATTCATTAACAATACT 60527  
 QY 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479  
 Db 60526 TTGATCATTTTGTATGTCGTATGTCGAAGATCGCAAGCGAACCTGTTTACCCCTATGCAT 60467  
 QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499  
 Db 60466 GTTTTAGTTAATGATGCCCTTCAAAAGAGCTTCAGCCGACGGTAAATGACACACCTACA 60407  
 QY 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAla 519  
 Db 60406 ACTCTATCAGTTTCTTAATGCTTTTGTATAGCGCGCTCCTCTGCTTTAATTTTGA 60347  
 QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
 Db 60346 TTTATTTATGACCATAGGATTCCTTAGGAATCTTATGCTCTCTTTATTTATGACCACT 60287  
 QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552  
 Db 60286 CTGTTGTTTGTATGCTCGTAAAGAAATTCCTCAAAA 60248  
 RESULT 4  
 LOCUS AP002547/c  
 DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,  
 section 3/4.  
 ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008  
 VERSION AP002547.2 GI:10176693  
 KEYWORDS Chlamydomophila pneumoniae J138  
 SOURCE Chlamydomophila pneumoniae J138  
 ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE 1  
 AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,

Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,  
 Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,  
 Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.  
 Comparison of outer membrane protein genes omp and pmp in the whole  
 genome sequences of Chlamydia pneumoniae isolates from Japan and  
 the United States  
 J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)  
 20298986  
 10839753  
 2  
 Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
 Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.  
 Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA  
 Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
 20330349  
 10871362  
 3 (bases 1 to 300550)  
 Shirai,M.  
 Direct Submission  
 Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University  
 School of Medicine, Department of Microbiology; 1-1-1  
 Minamikogushi, Ube, Yamaguchi 755-8505, Japan  
 (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, tel:81-836-22-2227,  
 Fax:81-836-22-2415)  
 On or before Sep 15, 2000 this sequence version replaced  
 gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.  
 gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.  
 Location/Qualifiers  
 1..300550  
 /organism="Chlamydomophila pneumoniae J138"  
 /mol\_type="genomic DNA"  
 /strain="J138"  
 /db\_xref="taxon:138677"  
 /note="synonym:Chlamydia pneumoniae (strain J138)"  
 /gene="dapF"  
 /complement(21..794)  
 /gene="dapF"  
 /complement(21..794)  
 /codon\_start=1  
 /transl\_table=11  
 /product="diaminopimelate epimerase"  
 /protein\_id="BAA98725.1"  
 /db\_xref="GI:8978890"  
 /translation="MAPYSPSTISKVFIYSGAGNRFLLGETLPEVEDVRFLOCBTRVD  
 GFVLKPSQCADQLIIFNSDGRPTMCGNGLRCAIAHLASOKGSDISVSTDGLYS  
 GYFVMDRVLVDMTLADRASVHLSRPDLPEKVVCIHTGVPHAVILPEISTDL  
 SILGPFURYQTFTSPGVNVNFVQLGHGCLRVTYERGVEGTAACTGALASALVV  
 SNSYGMKESIQTHTWGGELMTVSQNRGRVYLQGSVTRDL"  
 /complement(763..1338)  
 /gene="clpP\_1"  
 /complement(763..1338)  
 /gene="clpP\_1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CLP protease"  
 /protein\_id="BAA98726.1"  
 /db\_xref="GI:8978891"  
 /translation="MADGEVHKLRDIIIEKELLEARRVFFSEPTVEKASDAIKLWL  
 ELKDPGRPIVFINSPGSGVADGAVMDQIKMLTSPVITVTGLAASMGSVLSUCAAP  
 GRRFATPHSRIMIHQPSIGGPIITGQATDLDIHAREILKTIKARIIDVYVEATNQPRDI  
 EKAIIDRMWMTANEAKDFLLDGLILFSFNDL"  
 /complement(1358..2917)  
 /gene="glyA"  
 /complement(1358..2917)  
 /gene="glyA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="serine hydroxymethyltransferase"  
 /protein\_id="BAA98727.1"  
 /db\_xref="GI:8978892"  
 /translation="MLKVFEKFKFAIVEITKVVAVSLHLKPLENAGSKGOSLAS  
 TAYLAALDHLNAPPSIGERIDELKQSRHLKMIASENYSLSQVAMGNLLTDKYC  
 EGSPFKRFYSCCNVDAIEWECVETAKELFAADCAACVQPHSGADANLLAVMAILTHKV

gene  
CDS  
3107. .3778  
/gene="CPj0522"  
3107. .3778  
/gene="CPj0522"  
/codon\_start=1  
/transl\_table=11  
/product="CT433 hypothetical protein"  
/protein\_id="BAA98728.1"  
/db\_xref="GI:8978893"  
/translation="MTLYLGNOKTARKYQAHVLPILTLPPYAKSTPONKRALQFLPQ  
ATHVLLSPSSTHLFLSRMTSLLSKATLKTKYLCIGESTKERLLSLFGQVKYVATQ  
EIARGIFPLQALPSSARILYPHSSLARPIREFLNRFYFFSPYPIVYPRKUKNI  
LSKYKKIIF"SPSTVRAPAKIFPRPEKTYWQGRMTLQBFQKPSQKQVSLLETGK  
SRTSP"  
3843. .4175  
/gene="CPj0523"  
3843. .4175  
/gene="CPj0523"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98729.1"  
/db\_xref="GI:8978894"  
/translation="MASSATPGDGTAPSLFPPTATRPYRNFKALFVITIAIALVWIAL  
IATTAIGLCIHPLCSFIFUTAIPLFYISHYARNVYIALDVVPDHSKLQWRS  
HSPFSDR"  
4223. .5302  
/gene="CPj0524"  
4223. .5302  
/gene="CPj0524"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98730.1"  
/db\_xref="GI:8978895"  
/translation="MSGPSTESSQSVLSVFPDKEIAPKQFTIAKISTAILASL  
ALGAVAGISLTIIVGNPFLALDITALLSVTFVLYHQMTKSVSNWQKVLQNFK  
PLGKAWQKNDVQYNNHNPFRKVAIQTDASQFPQFTFTGLRVIENKST  
GIIPNVPGLINDNTNLSTILYSLTKDKSVMDCKQREGPAKEDPSPTEYRV  
VKLPNEALDOTNLNLSAEKSLTLPFLHGVCGPKSEELPNQOEYVROALLAVENCL  
KAIESHAALVALPLFTSVYVEPPEILPREGTFYWDNQIQAFCRKALLDAIQNTAIR  
YQKSLVILQDPFTIESQSRSEE"  
5642. .6406  
/gene="CPj0525"  
5642. .6406  
/gene="CPj0525"  
/codon\_start=1  
/transl\_table=11  
/product="CT398 hypothetical protein"  
/protein\_id="BAA98731.1"  
/db\_xref="GI:8978896"  
/translation="MHDALLSILAIQELDIKMLRMVKEKOKELAKVOSLKSDIER  
KVQKELEMENLKTQIRDGNRQIEISEQINKLENQQAAYKKMDEFNALTQEMTANK  
ERSLEHQLSDLMQKQAGEDIVSLKESLASENSVSVIEKEIFESIKKINEGKAL  
LEQRTLEKHTNPELLSIYERLLNNKDRVYVPIENRVYVSCGHVILTPQHENLVRKDD  
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAAV"  
complement(6827. .7816)  
/gene="kpsF"  
complement(6827. .7816)  
/gene="kpsR"  
/codon\_start=1  
/transl\_table=11  
/product="GutQ/kpsF family sugar-p isomerase"  
/protein\_id="BAA98732.1"  
/db\_xref="GI:8978897"  
/translation="MPFMIISTDVCQDILQKQKEAVDFFQAFQPKPEAMQLEKILGH

gene  
CDS  
3107. .3778  
/gene="CPj0522"  
3107. .3778  
/gene="CPj0522"  
/codon\_start=1  
/transl\_table=11  
/product="CT433 hypothetical protein"  
/protein\_id="BAA98728.1"  
/db\_xref="GI:8978893"  
/translation="MTLYLGNOKTARKYQAHVLPILTLPPYAKSTPONKRALQFLPQ  
ATHVLLSPSSTHLFLSRMTSLLSKATLKTKYLCIGESTKERLLSLFGQVKYVATQ  
EIARGIFPLQALPSSARILYPHSSLARPIREFLNRFYFFSPYPIVYPRKUKNI  
LSKYKKIIF"SPSTVRAPAKIFPRPEKTYWQGRMTLQBFQKPSQKQVSLLETGK  
SRTSP"  
3843. .4175  
/gene="CPj0523"  
3843. .4175  
/gene="CPj0523"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98729.1"  
/db\_xref="GI:8978894"  
/translation="MASSATPGDGTAPSLFPPTATRPYRNFKALFVITIAIALVWIAL  
IATTAIGLCIHPLCSFIFUTAIPLFYISHYARNVYIALDVVPDHSKLQWRS  
HSPFSDR"  
4223. .5302  
/gene="CPj0524"  
4223. .5302  
/gene="CPj0524"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98730.1"  
/db\_xref="GI:8978895"  
/translation="MSGPSTESSQSVLSVFPDKEIAPKQFTIAKISTAILASL  
ALGAVAGISLTIIVGNPFLALDITALLSVTFVLYHQMTKSVSNWQKVLQNFK  
PLGKAWQKNDVQYNNHNPFRKVAIQTDASQFPQFTFTGLRVIENKST  
GIIPNVPGLINDNTNLSTILYSLTKDKSVMDCKQREGPAKEDPSPTEYRV  
VKLPNEALDOTNLNLSAEKSLTLPFLHGVCGPKSEELPNQOEYVROALLAVENCL  
KAIESHAALVALPLFTSVYVEPPEILPREGTFYWDNQIQAFCRKALLDAIQNTAIR  
YQKSLVILQDPFTIESQSRSEE"  
5642. .6406  
/gene="CPj0525"  
5642. .6406  
/gene="CPj0525"  
/codon\_start=1  
/transl\_table=11  
/product="CT398 hypothetical protein"  
/protein\_id="BAA98731.1"  
/db\_xref="GI:8978896"  
/translation="MHDALLSILAIQELDIKMLRMVKEKOKELAKVOSLKSDIER  
KVQKELEMENLKTQIRDGNRQIEISEQINKLENQQAAYKKMDEFNALTQEMTANK  
ERSLEHQLSDLMQKQAGEDIVSLKESLASENSVSVIEKEIFESIKKINEGKAL  
LEQRTLEKHTNPELLSIYERLLNNKDRVYVPIENRVYVSCGHVILTPQHENLVRKDD  
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAAV"  
complement(6827. .7816)  
/gene="kpsF"  
complement(6827. .7816)  
/gene="kpsR"  
/codon\_start=1  
/transl\_table=11  
/product="GutQ/kpsF family sugar-p isomerase"  
/protein\_id="BAA98732.1"  
/db\_xref="GI:8978897"  
/translation="MPFMIISTDVCQDILQKQKEAVDFFQAFQPKPEAMQLEKILGH

SGWVPSGVKSCCVARKLVATLQSLSERALFSPVDLLHGDGLGLVSPGDIVCLFSSKS  
GETOELEDDVPHLKSRRALVAITSMPYNSLAASDLVILPVAELDPFLNLIPTNST  
TCQMIIFEDFLAMLLPFSRGVSLSTYGNHPSGQVMKANGKVKDFMPKTEVPFCHLG  
DKVSLFEGVSGVACGCVCTVDPQRLMGIFTDGLRRSLASYGVEVLSLSLRKVMTA  
NPRCITEDSDIAIALQLMESSSPVAVLPVLDNEENRHVTGLLHMHHTLAKAGLL"  
complement(7843. .9024)  
/gene="sucB\_2"  
complement(7843. .9024)  
/gene="sucB\_2"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrolipoamide succinyltransferase"  
/protein\_id="BAA98733.1"  
/db\_xref="GI:8978898"  
/translation="MIFEPFRPKIGETSSGGSIVRWLKNLGDHVARDEPLIEVSTDKI  
ATLPSPKAGRLVPCNEGDEVASGDVGLGLEIELEISEADDESTCPLTSCCTKSEA  
GSSSSWFSPPASVLSLAQREGIGLDNLQIAGTKGGRVTRQPLEAVISQOVSPIPE  
IFQGVNRPMSPLRAIASLSSKSSDEVPHASLVVDVDTLMLNISGRQRFLDTH  
GVKLITSTIVQCLAQTLRQFPLNGSLDGTTLVMKKSNNVGVANLNKREGVIVRYPEV  
NCDRLGVSIAKALADLSRLNKLDPSVDGVSVTNFGMTGALIGMPIIRYPEV  
AIIIGICTIQKRVVVRDDDSLAIKRMVYVTLTDFHRLVDLGIYGSEFLTSLKRLSEVTM  
G"  
complement(9038. .10282)  
/gene="gitT"  
complement(9038. .10282)  
/gene="gitT"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate symport"  
/protein\_id="BAA98734.1"  
/db\_xref="GI:8978899"  
/translation="MKLWMKIFIGLVGVTLGLVEDKAIFFKPIGDIIFLNLISMVYV  
PIVFCMSVILGIASIDMKKLGRIKSVGLVIGTTALAVIGLCFAMIFSPGNGCDFPA  
OQOSMDSAVTVIDSNKTAAVFLSIIAQVFPSPNVPVSFAEGNILOIIFIIFLGLALRL  
SGERGPVERFIDGPEIIMLVNMTMSPAP'GVGASMAWISGNHGLVWLQKGRPII  
AYLACLPHATLVFGGLVRFCKMSFKSLSNMDAISCAVSTASSATLPTVMRCVS  
KNLGSAEVSQVFLPLGATVNMNGTAIFQGMMAVFAQAYNCPLSUSSLLLLVVTATF  
SASGAGVPGGMITLGSVLASVGLPIQIAI:LAGIDRLRLDIVGTMPNLLGDAVATY  
VASGEGELSPYESIKQESVETT"  
complement(10282. .11379)  
/gene="ycaH"  
complement(10282. .11379)  
/gene="ycaH"  
Alignment Scores:  
Pred. No.: 4,19e-237 Length: 300550  
Score: 2730.00 Matches: 551  
Percent Similarity: 99.28% Conservative: 0  
Best Local Similarity: 99.28% Mismatches: 1  
Query Match: 99.27% Indels: 3  
DB: 1 Gaps: 0  
US-09-868-987-14 (1-552) x AP002547 (1-300550)  
Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 50924 ATGGTCAGCAGCCCTATTTAAACGTCCTCCATGAAAATCATGCCAGTGTCTCAGGAAA 50865  
Qy 21 PreThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 50864 TTTACCCACCGTGAAGTACGAAACATCGCTCAGATTTTAAATCTGGAGCGATGCTCTTT 50805  
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
Db 50804 GTTCCCGAGGTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACA 50745  
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 50744 CAAGGCATTATCTCAGCATGCTGTGGCAATGCTTATTCTTTTGTATGAGCGTATAT 50685  
Qy 81 TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr 100  
Db 50684 TATAGATTGGAGGCGTCATCGCTTCGGAGCTGTCTTCTGAATCTTTTGTCTATCTGG 50625

```
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 50624 CGAGCTCTACAGTATTGGATGGCGCACTACCTTGTCCAGNCTCGCTGGGATGTTCTT 50565
Qy 121 AlaMetGlyMetAlaValAspAlaSerValLeuValPheGluArgGluGluPhe 140
Db 50564 GCTATGGGATGGCGCGTAGATGCAATGTTCTTGTATTGCAAGAATCCGAGAGGAATTT 50505
Qy 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 50504 TTAATTGCTCTCAAGGCTTAAATAAACTGTAGAAAAAAGGATATACCAAGGCTTTTGGAGCC 50445
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 50444 ATTTTGTATTCTAACTTCACTACAGTATTGGCTCAGCACCTCTTTTCTTCTAGATACA 50385
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Db 50384 GGGCCTATTAAAGGGTTTGGCTTTGACATTGATTTTGGAAATTTTCTCTTCAATGTTTACG 50325
Qy 201 AlaLeuPheMetThrLysPhePhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220
Db 50324 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAGACCCCAACATACACAG 50265
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db 50264 TTGCATATGATGAATAAGTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCAGAAAA 50205
Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 50204 CTTTGGGCTGTTTCTGGAAGTGTTTCTTTAGTGTGCTGTCTCGGGTTTGGAGCC 50145
Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280
Db 50144 TGGAAATTCGTTTGGGAATGATTTTAAAGAGGGTATGCTTTTACCTTTTAATCCAAAA 50085
Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db 50084 GAGCATGGCATCAGCGATGTTGCTCAATGCGTGCAAAAGTTGTGCATAACTACAGAA 50025
Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320
Db 50024 GCTGCTCTTCTCTAGAGACTTCGCTATCAACATTTGGATCTTCAGAAAAAGATCAA 49965
Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340
Db 49964 ATCTATTTTAGTGATAAAGCTTTAAGCTATATACTAAAGCAGATACGAGGCTCTCTCTAAA 49905
Qy 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360
Db 49904 ATTAACGATCATGACTGCGGTAGCTGTGGATTTGTGCAGAAACAGGCTAGATTTC 49845
Qy 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379
Db 49844 TCTACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAAGGTAAAGCAGCAAACTATCG 49785
Qy 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399
Db 49784 AAGAAATTCGCTTATCAGCGCACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 49725
Qy 400 TyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419
Db 49724 TATGTAGTTTGGCTTTTGAATGGCAATATGCTTTTCAGTGGCGTATGCGCTTTTAATTCAT 49665
Qy 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysIleGln 439
Db 49664 GACCTTTTGGCTTACCTGTGCACTCTCTTTATAGCACATTTCTTTTCAAGAAAAATTCAA 49605
Qy 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThr 459
Db 49604 ATAGATTTGCAAGCAATTTGGTCTTTTAATGACTGTATTGGGTATTTCATTAACAATACT 49545
```

```
Qy 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db 49544 TTGATCATTTTGTATCGTATTCTGTAAGATCGCAAGCAACCTGTTTACCCCTATGCAT 49485
Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db 49484 GTTTTAGTAAATGATGCCCTTCAAAGACGTTTACGCGCACGGTAATGACAACAGCTACA 49425
Qy 500 ThrLeuSerValLeuLeuMetLeuPheIleGlyThrLysSerValPheAsnPheAla 519
Db 49424 ACTCTACAGTTTGTAAATGCTTTGTATAGCGGCTCTCTGTCTTTTAATTTTGA 49365
Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db 49364 TTTATTATGACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTTATATGACCACT 49305
Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db 49304 CTGTTGTTGTTTATGTCGTAAGAAAAATCGCTCAAAA 49266
RESULT 5
AE016994 304769 bp DNA linear BCT 14-APR-2003
LOCUS Chlamydomophila caviae GPIC section 1 of 4 of the complete genome.
DEFINITION AE016994 AE015925
ACCESSION AE016994.1 GI:29834116
VERSION
KEYWORDS
SOURCE Chlamydomophila caviae GPIC
ORGANISM Chlamydomophila caviae GPIC
REFERENCE Chlamydomophila caviae GPIC
1 (bases 1 to 304769)
AUTHORS Read T.D., Myers G.S., Brunham R.C., Nelson W.C., Paulsen I.T.,
Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
Umayam L.A., Haft D., Peterson J., Beanan M., White O., Salzberg S.,
Salzberg S.L., Hsiao R.C., McClarty G., Rank R.G., Bavoil P.M. and
Fraser C.M.
TITLE Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
examining the role of niche-specific genes in the evolution of the
Chlamydiae
JOURNAL Nucleic Acids Res. 31 (8), 2134-2147 (2003)
MEDLINE 22569155
PUBMED 12682364
REFERENCE 2 (bases 1 to 304769)
AUTHORS Read T., Myers G., Brunham R., Nelson W., Paulsen I.,
Heidelberg J., Holtzapple E., Khouri H., Federova N., Carty H.,
Umayam L., Haft D., Peterson J., Beanan M., White O., Salzberg S.,
Hsiao R.C., McClarty G., Rank R., Bavoil P. and Fraser C.
Direct Submission
TITLE Submitted (29-OCT-2002) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..304769
/organism="Chlamydomophila caviae GPIC"
/mol_type="genomic DNA"
/isolate="GPIC"
/db_xref="taxon:227941"
57..1052
/gene="hemb"
/locus_tag="CCA00001"
57..1052
/gene="hemb"
/locus_tag="CCA00001"
/notes="similar to GB:L16784, SP:P33395, and PID:290384;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="delta-aminolevulinic acid dehydratase"
/protein_id="AA04754.1"
/db_xref="GI:29834117"
/transl_table="MSSLALLRRPRNRKTAIRDLVAETALLPQDFICPFFVKEGKN
ICEIESLPGVYRWSIDLKLEIERLCSGLKXAVILFPVPSDLKDAYGVSSNPKN
LCKSIYEVKAFDPDLCVISDIALDPYTHGDGIMDOCEVINDSESVIRFGNIATLHAE
MGADIVAPSDMDGRVAHIRAKLDQSGWTNTLILSYSVKVASILYNFRDLGSLHQS
```



/product="exodeoxyribonuclease V, gamma subunit"  
 /protein\_id="AAP04761.1"  
 /db\_xref="GI:29834124"

## Alignment Scores:

Pred. No.: 2,03e-178 Length: 304769  
 Score: 2088.50 Matches: 421  
 Percent Similarity: 83.81% Conservativity: 50  
 Best Local Similarity: 74.91% Mismatches: 72  
 Query Match: 75.18% Indels: 19  
 DB: 1 Gaps: 3

US-09-868-987-14 (1-552) x AE0016994 (1-304769)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
 Db 189026 GTGGTCAGCGATCTCTGTTAAATGCTCTTTAAAGATACGCTAGTGTTCTTGGAAAA 189085  
 Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
 Db 189086 TTTTCACATCGTGAAGTAACCGCTAGCTACGGATTAAATCAGGATCTATGTCCTTT 189145  
 Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
 Db 189146 GTTCTGAGGTTTTAAAGTGAAGATTATCTCTCTGAATTAGGCAACAAACGATCA 189205  
 Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
 Db 189206 CAAGGATTATCTCTATATGTTAGTCTCTGTTTGTATGATCAATTTTGTAGCGCTTTAT 189265  
 Qy 81 TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuLeuAsnLeuLeuIleTyr 100  
 Db 189266 TACAAATTCGAGGAGTTATTGCTCGGAGCTGTTATCTTAAATCTTCTATTGATTGG 189325  
 Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 189326 GCAGCCCTACAGTATCTTGATGCCCTCTAATCTCTACGGGATTAGCGGGAATTGTATTG 189385  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgGlyLeuGluPhe 140  
 Db 189386 CCATGGGAATGGCTGTAGATGCTNACGTGTTGTTTTCGAAGAATTCGGCAAGAGTAT 189445  
 Qy 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 189446 TTATTATCGGAAGCTTACCAGCTCTGTAAGACGAGGTATAAAAAGCTTTCAGCGCT 189505  
 Qy 161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
 Db 189506 ATCTTTGATTGCAATTTTGCACCAATATTAGCTTCGCTACTTCTTTGGTTTATAGATACG 189565  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 189566 GGACCAATTAAGGAATTCGCTGACTTTAATCCTAGGATTTTCTCTTCATGTTTACT 189625  
 Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220  
 Db 189626 GCATTATTATGACGAGATTTTCTTCATGATGATGAATAAACCAGCAAGAACTCAG 189685  
 Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
 Db 189686 CTGCATATGATGAATAAGTTTATTCGGCATCAAGCATGATTTCTTAAAGAGTGCAGAAAGA 189745  
 Qy 241 LeuTyrAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 189746 CTTTGGATGGTATCCGGGAGGCTTATTGTTTGGGATCGGTGCGTTAGGTTTCGGGGCA 189805  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280  
 Db 189806 TGGAAATTCGTTTGGGAATGGAATTTTAAAGTGTGTACGCTTTCACATTTAGACATGCT 189865  
 Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
 Db 189866 GATCAGAAAGCTGTAGATGTTGTACAGTTCCTGTAATAAGTTAGGGGATAAATTTAAGCAG 189925

Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
 Db 189926 GTAGGCTTATCTCTAGAGACTTCAAAATTAAGACGTTAGACTCTTCTGAGAGATAAAA 189985  
 Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSer----- 337  
 Db 189986 ATATACTTCACTCAAAACGCAATTAAC-----CGAGTACAAGCTTCTGAGAAACAG 190036  
 Qy 338 -----LeuLeuLysLeuThrIleMetSerTyrPheGlyCysGly 350  
 Db 190037 GTATTGAAACCCCTGATCTTAATTTATCTTATGATGGCATACTTCTTCATACAGGA 190096  
 Qy 351 IleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPhe 370  
 Db 190097 ATAGACACTTCTTCAGAG-----AGCTTAAAGATACCAGCAATTTTC 190138  
 Qy 371 TrpSerLysValSerSerLysLeuSerLysMetArgTyrGlnAlaThrIleGlyLeu 390  
 Db 190139 TGGCTTAAAGTTAGCGACAGTTCTCTAATAAGATGCTCAACAAGCTTGTATAGCGTTG 190198  
 Qy 391 LeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTyrGlnTyrAla 410  
 Db 190199 GCGGAGCCCTTAGCGATCTTCTTATGTAGCTTGCCTGCTACCTGCGCGCTATTAGTAGCC 190258  
 Qy 411 PheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIle 430  
 Db 190259 TTTAGTCTGCTTTCGCGATTTGATTCATGATCTTGTGCTACCTGCGCGCTATTAGTAGCC 190318  
 Qy 431 AlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThr 450  
 Db 190319 ACACATTTCTTTTGCAGAAAGATACAAATGATTATCAAGCAGTCGGAGCTTTGATGAGC 190378  
 Qy 451 ValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAspArg 470  
 Db 190379 GTTTTAGCGTATTCGTTAAATAATACCTTAATCATCTTTCGATCGTATTCGAGAAGATCGT 190438  
 Qy 471 GlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPhe 490  
 Db 190439 CAGGAAAAATTTATTTACCTATGCGGATTTTAAATAACGAGCAGCTGCAGAAAAACCTTA 190498  
 Qy 491 SerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuPheIle 510  
 Db 190499 GGAAGAACAGTAAATGACACAGCGACACACACTGTCGGTATTATTAATTTTACTGTTGTT 190558  
 Qy 511 GlyClySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThr 530  
 Db 190559 GGAGCGCGTTCAATCTTTAACTTTGCTTTCATCATGACAATAGGTATTCTTTTGGGCACA 190618  
 Qy 531 LeuSerSerLeuTyrIleAlaProLeuLeuLeuPheMetValArgLysGluAsnArg 550  
 Db 190619 CTATCATCTCTATACATAGCACCCCTCTCTTTTATTCATGGTTCTGTAAGAGGAAAGA 190678  
 Qy 551 SerLys 552  
 Db 190679 ACAGG 190684

RESULT 6  
 AE001318/c  
 LOCUS  
 DEFINITION Chlamydia trachomatis section 45 of 87 of the complete genome.  
 ACCESSION AE001318 AE001273  
 VERSION AE001318.1 GI:3328875  
 KEYWORDS  
 SOURCE  
 ORGANISM Chlamydia trachomatis  
 Chlamydia trachomatis  
 Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE  
 1 (bases 1 to 989)  
 AUTHORS Stephens,R.S., Kalkan,S., Lammel,C.J., Fan,J., Marathe,R.,  
 Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
 Koonin,E.V. and Davis,R.W.  
 Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis











SINKGDIVIVTDHHTPTGKLPCHCIATLNPKLDKNSYPNKELTGVGVAFKLVCAATYEE  
 LIKQASWKIDILHFLDVLSTGIADVRLSGENRILVSYGKEITKGKRLGLKKL  
 CTSQVSDKSNLGRITPKNLSNGLRSLDSSQGVKLLSLQSDPRVTGTVLSLAV  
 NOERQTEAEVLRVERITLAPLANPKLTQASAIVLASPVWHSRVIPILISARLARTYKPV  
 AIALQDGLCKSRRTIGAPLKLGRKCESFFLSYGGHDFPAAGLMKEIQEIGFRKK  
 FHLVNSLRKDEATILPLDVCMDFFSOISODLSSMELLEPPFGKGNPSVFFYTRAIQ  
 VRYPLKLSGNHVKULYLSNGENRLEGTAFQGGQDKISLIKANNMPLDIATYLRVKNRSV  
 HGAVRLIQDFRIQAP"

gene complement (8009..12211)

CDS /locus tag="TC0733"

complement (8009..12211)

/locus\_tag="TC0733"  
 /note="similar to sp:P07114, GB:S96966, GB:X04830, and  
 PID:42533; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=1

/product="secDF protein, putative"

/protein\_id="AAF39543.1"

/db\_xref="GI:7190762"

/translation="MKRNFRNLSVIFVFALALYHLVPTCLYYSRPLNKKIEGREAH  
 KIIRLTOQVADARNDLTVRKRLSTLKLKRGSKVPHLSIPGVNVHFPPLAEDAAVFI  
 DNVVHGHSVPTKARLYVLGSSSLTGETIIVOVSGSLVTSLTEDFSFVSDKLVV  
 SOELENTATNLPVNVESCDHGYSLSWDSPIEKVAHARGICSLGLRLTEKTOAEL  
 SRFGTQDFVFALSRVENALHEDVKEYREELNEAAYLLHNRNLNRWNSPRVVGNV  
 VDCSELSPCSSMSLTGGKLLFHDADVIARQQLVGDRLDLERLFAVEKQHIATR  
 LNRVETESGFARLDEAASGKIVLQGHRCVORIVEHLTALVLRPIAETCDLSIE  
 NFPVSRPEMNDTLGCFIFSPDSSCRHSKGVYIVFKGLRSIIAKCESSGPOEAGL  
 LQDOLNLYACFIHTDAVMSLSLGRDLVLEIKELPQVIOVWGEFFVQSGEASLEVRD  
 VRDLAVVNRLEKQAQELVNDWDDQYRQAQCSMNPQVRLRAAIPHKNVFENFKLNV

#### Alignment Scores:

Pred. No.: 3.31e-165 Length: 12442  
 Score: 1924.00 Matches: 384  
 Percent Similarity: 79.30% Conservative: 72  
 Best Local Similarity: 66.78% Mismatches: 73  
 Query Match: 69.26% Indels: 46  
 DB: 1 Gaps: 5

US-09-868-987-14 (1-552) x AE002341 (1-12442)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
 Db 9679 GTAATTAGTATCCCGTGTAAATGTTCCGCTCAAGATCATGCAAGTGTTCCTGGGAAT 9620  
 Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
 Db 9619 TTTTCGTACAGGGAAGTCAATCGTTAGCTTCTGATCTTAAATCTGGGGCTATGTCCTTT 9560  
 Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
 Db 9559 ATCCCTGAAATTTTAAAGTGAGGAAGTAGTTTCTCCAGAGATTAGCACACTCTCAGAGGTTG 9500  
 Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
 Db 9499 CAAGGGATTCTCTCTGATTTTAAAGTACTAGTGGTCTCTCATTTGCAATTAATCAGTGTGTTAT 9440  
 Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100  
 Db 9439 TACAAGTTGGGGGTGTCATTCCTCGGACCGCTCTTCTTAATTACTACTTATCTCGG 9380  
 Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 9379 GCATCTATGCAGTATTAGATGCTCTCTTACTCTATCAGGACTAGCTGGAAATATCTTA 9320  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArGileArgGluGluPhe 140  
 Db 9319 GCAATGGGGATGGCGGTTGATGCTAACGTCCTGGTCTTGTGAAGGATTTCGAGAAGAATAT 9260  
 Qy 141 LeuLeuSerGlnSerLeuLysSerValGlnLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 9259 TTATTGAGTAGAAGTTATACAGATCTGTAGAGGCTGTTATTAAGAGGCTTTTAGTGCA 9200  
 Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180

Db 9199 ATTTTTCATTCTAATTAAACAATATTTCGCTTCGCAATGCTTCTCATGTTGGATACA 9140  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 9139 GGACCAATTAAGAGGTTGCTTTGACCTCATCATCGAAATTTTTCTTCTATGTTTACA 9080  
 Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
 Db 9079 TCITTTGTTTCATGACGAAGTTCTTTTCTTCTGATACAAAAACAGAGACCCAA 9020  
 Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
 Db 9019 TTACATATGATGAACAAGTTTCATAGGGGTCAACATATAATTTCTTGAAGAATGTAAGCGT 8960  
 Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 8959 TTGTGGGTGTTTTCCGGGGCTGTAGTCGTTTATAGTTGTATTGGGCTTAGGCTTTGGTGCC 8900  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyLysTyrAlaPheThrPheAsnProLys 280  
 Db 8899 TGGGACTCCATTTTAGGTATGGATTTCACGGGAGGATACCGCTGACCTTAGAT----- 8846  
 Qy 281 GluHisGlyIleSerAspValAla-----GlnMetArgGlyLysValVal 295  
 Db 8845 -----TCTGATGTTTGTGCCTATAATCCACATCAGATGTGTTCTGCTTTAAAA 8798  
 Qy 296 HisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySer 315  
 Db 8797 AAACGCTTCGGGAAATTTGGGTTGTTCTCTCGTATTATCTGCTGTAAGCTGATAGC 8738  
 Qy 316 SerGluLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArg 335  
 Db 8737 TCTGGAAGAAGTTAAGATCTATTCTTCTCAAAATGCTTTAGCTCGTGTGACCAAGCCAAA 8678  
 Qy 336 -----AlaSerLeuLeuLysLeuThr 342  
 Db 8677 GGGGGGTATTCGGAACAGAATGGGCGGACCATCATCTAGCTCAAGTTTTCGAAGTATTA 8618  
 Qy 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362  
 Db 8617 TCCGATTCT-----AGT 8606  
 Qy 363 AsnSerLysArgAsnAlaLysPhe-----TrpSerLysValSerSer 376  
 Db 8605 GATGCTACTTCTCTCGCAGCATTTGAATCTCTCGAGGAAGTTGTTTAAAGTTAAGTGA 8546  
 Qy 377 LysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle 396  
 Db 8545 CAGCTTTTCAAAATGCAACACAGGCAATCATGCAATATTGCGAGCTTTTAGGAATT 8486  
 Qy 397 IleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAla 416  
 Db 8485 ATCTTACTTTATGTTAGCTTGGGTTTTGAATGGAGATATGCTTTTCAGTGCATTTTGCT 8426  
 Qy 417 LeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLys 436  
 Db 8425 TTAATGCAATGTTTGTAGGACCTGCTCTGTATTAGTCTCTGCTCTGCTATTTCTTTTCAT 8366  
 Qy 437 LysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeu 456  
 Db 8365 AGAATACAGATTGCTTACAAAGCAGTGGGGCGCATTTGATACGGTGTGTTAGGCTACTCTTTG 8306  
 Qy 457 AsnAsnThrLeuIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThr 476  
 Db 8305 AATAATACGTTGATTTATTTTGTATGCTGTTATTCGAGAGACCGAGCTGAGAAATTTGTTCCACC 8246  
 Qy 477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496  
 Db 8245 CCAATGCCCATATGATCAATGATGCTTTTCAGAAGACTCTGGACGCTAGCGTCAATGACA 8186  
 Qy 497 ThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPhe 516



VLIRIQADGGENAEQSAITLVRGELEDKDYDFRRVVEVGVPAVSGDLTVTSTIGVLLAM  
AAIIVYIVRREWFQFALGAVISMVHDVVFTIGLFLVGLIEFNLSIAAILTIIGVSLN  
DTVIYDVRNEMRLRYKMPISMIIDVSLNOTLGRITLGTLLALLALALVLFGEVI  
RSFTFAMLFQGVIGVSSVYIAAPVLIAFKLRPSKDAEDKDKKENTNTGAIQKPA  
V"

## gene

/complement(5493..5840)

/gene="yajc"

/note="synonym: Atul563"

/complement(5493..5840)

## CDS

/gene="yajc"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="preprotein tranlocase protein"

/protein\_id="AAL42567.1"

/db\_xref="GI:17739991"

/translation="MFFSQAFQAADGASGFGSGLEMLFLFAPLMVVMVFFLIRPOR  
AQMRRQETLTSIRRGDQVILGGVIGVTKYVIDDNELEVEIAEGVKIRAAAYIAEV  
RVKGEVKTAAAS"

## gene

/complement(5493..5840)

/gene="Atul564"

/note="synonym: Atul564"

/complement(5493..5840)

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAL42568.1"

/db\_xref="GI:17739992"

/translation="MIRETASLLLSLRLLTNAVERLAGPAPAAANDWNAADCFWAP  
LNSFLQVPRNRILKILGVHDVDRDILHENTLRFAGFAANNVLLWARGMKSSL  
KXVHEDVRAAGSUSGLKVEVREDIHTLPALLDLKISDQRFILFCDLSPDHDH  
YKSLAALDQEGPDNVLVYATSNRRHLPRHMMENESTAINPSAEVKEKVSLS  
RFLWGLFHKCSQEDYLMIDSVAEHFDGLDRLDTHLHAEALWATTGRSGRVAWQY  
IODAGRIRKPLDR"

## gene

/complement(5493..5840)

/gene="Atul565"

/note="synonym: Atul565"

/complement(5493..5840)

## CDS

/gene="Atul565"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="serine acetyltransferase"

/protein\_id="AAL42569.1"

/db\_xref="GI:17739993"

/translation="MHCVFLEFNMSVFKKIVFALSAPFVAPHIFLLMSSRHDNPIFRD  
LEVWSEKLLGKIKSGSLVASVVELLSLAPEFNLFYRVPRARFFKELSRPENTLH  
IHTRIGRPFQIRGFATIIIAESIGACDWINOQVTIGYVDSGTAPKIGDNVTINAGA  
KVLGNIVIGDNKAVGANAVVKNIPADCVVGVPAIIRRAGARVDEAL"

## gene

/complement(5493..5840)

/gene="Atul566"

/note="synonym: Atul566"

/complement(5493..5840)

## CDS

/gene="Atul566"

/note="identified by sequence similarity; putative; ORF

located using Glimmer"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAL42570.1"

/db\_xref="GI:17739994"

/translation="MSFGNTRRVVATTNRNPTYSFFTOQTAKFMIIICLDNRNRLPAG  
ITQSSKEAWEEWRPLFA"

## gene

/complement(5493..5840)

/gene="Atul567"

/note="synonym: Atul567"

/complement(5493..5840)

## Alignment Scores:

Pred. No.:	Length:	Matches:
4,068-64	10046	187
Score:	817.50	Conservative: 139
Percent Similarity:	58.42%	Mismatches: 190
Best Local Similarity:	33.51%	Indels: 43
Query Match:	29.43%	

US-09-868-987-14 (1-552) x AE009114 (1-10046)

Qy 1 MetValSerSerProLeuLeuAsnValPro---LeuLysAsnHisAlaSerValSerGly 19  
Dd 4477 GTTATTTCCGGCGCTGTATCCCGAGCCCATCATCGCGGTTCGGGACAGATTTCGGGT 4418  
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39  
Dd 4417 AGTTTCTCGGTGCGAGGTGCGAATGATCTCGGTTCTGCTCGGAGCGGTGCACTCCG 4358  
Qy 40 PheValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCys 59  
Dd 4357 GCCACGCTAACGGTTCGTTGAAGAACGACCGTCGGTCCGAGCTCGGTAACTCCATC 4298  
Qy 60 ThrGlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerVal 79  
Dd 4297 ACCGCGGCGCTTACGGCCAGCGCCATCGGCGCGTTCGCGTCTCTCATCTTTCATGTC 4238  
Qy 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuAsnLeuLeu 99  
Dd 4237 TTCTACCGCTTCTTCGGCGCTGCTGGCGAATATCGCGCTTATCGTCAACGTCGTAATGTTG 4178  
Qy 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119  
Dd 4177 ATCGCGGTGTTCGCGTCATCGGTCACCGTCACTTTCGCGGTATCGCGGTATCGTC 4118  
Qy 120 LeuAlaMetGlyMetAlaValAlaAsnValLeuValPheGluArgIleArgGluGlu 139  
Dd 4117 TTGACGATCGTATGCGGTGGATTCCACGCTTCTGATCTATGAGCGAATACCGGAAGAG 4058  
Qy 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159  
Dd 4057 GTGAAGACGCGCAACCGCTGATCCAGCGCTGGTAATGTTTTTACCGCGCTTTCGCG 3998  
Qy 160 AlaIlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAsp 179  
Dd 3997 ACCATCATCGACGCCAACCTGACGACCGCTGATCGTGGGAGCGGTGCTTCTACATCGGC 3938  
Qy 180 ThrGlyProLeuLysGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPhe 199  
Dd 3937 ACTGTCGCGTCAAGGCTTCGCGTCACGCTTCGCGTTCGATCATTACCACCGTGTTC 3878  
Qy 200 ThrAlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThr 219  
Dd 3877 ACCGCTTATACGCTGACGCGCTGGATGTTCCGCTACGCTGCGGTCGCGCGCGCTCCGAAA 3818  
Qy 220 GlnLeuHis-----MetMetAsnLysPheValGlyIleLysHisAspPheLeuArg 236  
Dd 3817 CATCTGCGAAGGTGTACGACCGCCATGTTTCGATGTCGCGATATTCCTTTCATCGGT 3758  
Qy 237 GlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGlyCysAlaLeu 256  
Dd 3757 TATCGCGCGGTGTTTTCATGATCACCCTG---GTGATCATGCTCGCTCGCTCGCG--- 3704  
Qy 257 GlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThr 276  
Dd 3703 GGTTTCGTCGCCAAGGCTCTGAACCTCGGCATCATTTCCAGGTCGCTCGGTTCATCGAA 3644  
Qy 277 PheAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValHis 296  
Dd 3643 GTGCGCGCCAGCAG---GGTGGAGCAGATCTTCTGACATCCG-----GAG 3599  
Qy 297 LysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSer 316  
Dd 3598 CGTCTGAACACGCTCAACCTCGGT-----GAAATCCAGGCTCAGAACTTCGCTCGCG 3545  
Qy 317 GluLysIleLysIle-TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAl 336  
Dd 3544 CAGGATGTGCTCATTCGCGCATTCAGCGCAGCGCGCGGCGGAAATGCGGACGCTCGGCC 3485  
Qy 336 aSerLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnAr 356











```

KKIRAEKDQKQERFVSGAVERGVSKPQADFI FDLAKFADYGFNKSAAAVAVVSQYOT
AYLKAHYVEVEFLAASMTDGMNTDKLADROQDALRGLIEVLAPSVMTSFRFPEVGENR
IYYSMAALKGVGDAVAEHI VAVRGEPFKNLADFCERVDPKI VGRVPESILMAGALD
CFGHRAACOMAGVERMGLASLAQONAVSGOAGDFGASLGQSQALNLPATDPMLAAD
RLHREFQVGVFSLSHPLDEYKAALOKMRVQNWAEFFSAVRKSGASAGLACTVTSKQB
RKRTIRGNKMGVVAESDTSQGEAVLFSALAOYRLDEAGRSVVITVAEDRPEGVNL
RINSVQSLEESAIRIQKALRIFVRNEGPAISIIQSOLTORGESQVSIIIVIKEEAQGEVE
IALOGGYRVSQIASAMRAVPGVVEVELV"
/gene="ml10871"
/complement (4811..6823)
/complement (4811..6823)
/gene="ml10871"
/codon_start=1
/transl_table=11
/product="propionyl-CoA carboxylase alpha chain precursor"
/protein_id="BAB48367.1"
/db_xref="GI:14021756"
/translation="MFTKILIANRGEIACRVIKTARKMGIATVAVSDADRAVHVEH
ADEAVHIGPSPAAQSYLPEKIIIACKETGAQVHPGYGFLSERAACEALEKEGIVF
IGPKPAIKAMGDKIESKFAANAARVSTVPVGLGVENADHAEIKAGEIGYPMVIKAS
AGGGKGMRIANWQSEVRDGFDRARSEAKSFGDQDVPIEKVVDPRIHIEYQVLADAH
GNALVIGRECSIORNQKVAEEAPSPFLDAKTRKAMGEQSVALARAVDYSAGTVEF
IYDKDKNYFLEMNTRLOVHPVTELVTDGLVEOMIRVAAGEKLSQSDIKLNGWA
VESRUYAEDPYNFLPSIGRLTRYRPPPEQFGDVIYRNDTGVTGEGSEISMFYPMILA
KLCTWAPTRI EAI DAMESVLDS FVVDGIEHNI PFLAALMQHPRWREGAISTGFLAEY
PDGFAPVPSNREKRAVLASI VTAVELLRDRDLRLGRLGAPHSGLVKRWDAVVRIGDDY
LQASILEGMSI PMEIDLSIDGKALT VSSDRPQDLIRWCTVGRKSVQAQIRVPANG
LRIANKGMSVTARAMPLPRIASLERLMPEKVPADTSSKLLLCMPGLGVLSIATVEQEVK
AGETLVAVEAMKNVLRARDELVVVSKLNAPGDSLAVDAVIMBEFA"
7023..8150
/gene="mlr0872"
7023..8150
/gene="mlr0872"
/codon_start=1
/transl_table=11
/product="glutathione dependent formaldehyde
dehydrogenase"
/protein_id="BAB48368.1"
/db_xref="GI:14021757"
/translation="MKTRAAVAVAAKGKLEIMEVDLDPREGEVLVEIKATGICHTDE
FTLSCADEGEPFALIPALHGEHGVVDVVGKGTYSVKGDHVIPLYPTECRQCPSCLSRK
TNLCATIRATQGGIMPDSRSRFSIGDKDI FHYMGCSFTSFNPLVPEITALAKVNPDP
FDKICYCGVTGIGTAVINTAKVEQGA TAVVFGLGIGLNVII QGLKLAGADMIIIGVD
LNNDKAMGERFTGFTFVNPKEIDGDI VPHLVNMTKRGADQI GGADYTFDCTGNTKYM
ROALEASHRGKSVVIGVAGAGOEI STRPQLATGRTWMGTA FGGAGRDRDVRIVD
WYMDGKIQIDPMI THLLKLEIDINKGFDLMHEGKSIRSVVY"
8113..8787
/gene="mlr0874"
8113..8787
/gene="mlr0874"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48369.1"
/db_xref="GI:14021758"
/translation="WRASRSVASFTEASHTPPTDRPATSVVQNWEEKRAEKLHPRI
DNGLPKESAGKGLTCACTSKPVKVKQGLAHNHACGCTCKWKPEGALFSVVAVA
GTGVDVTENSGDKLVVDASALIQRHACTGCGVHRHGPVERDHAFKGLTFIHPERFVE
DWSGPPGFTAFVSSIIYESGVDPKRMGDGIRQAIRLTIGUEPYDCLNPLGMDYATWTAKK
SGALPA"
8818..9318
/gene="mlr0875"
8818..9318
/gene="mlr0875"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48370.1"
/db_xref="GI:14021759"
/translation="MFPAPVRFGIEMPAFTIYVDADCPVKAEEVKEKVAERLVGVVTVV
SNGGLRPSRDMIRVNVVSKGADAADWIVENAKSNVNVVTADIPLAARAVALLGAHV
GPTGRPFPTTIGMAVMRDLKQLHRETGESKGNAGSAPQDRSRFLGELDRILRRAL
KSATPG"

```

```

gene
CDS

```

```

9338..10027
/gene="mlr0876"
9338..10027
/gene="mlr0876"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48371.1"
/db_xref="GI:14021760"
/translation="METAMAAETPIPRPVQRHMLFAMSACFGALAFVALLLHAPLAY
SIGANAPFAAYVILVVSQMPKFTSKYLSKNARATDQVLVIFAVTLVVVVIVAVVLSLEL
LINQKRGNSVEIGFALLSIPLCWFTIHAMAALHYAHVYMMDGDADAETFKKIPVGG
LLFPDGKRPGEWDELFSAVIGMTAQTADTNI STTHMRCVVLVHSILSFFFTVIVAA
AVNLAVSLGGP"
10125..11006
/gene="mlr0877"
10125..11006
/gene="mlr0877"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48372.1"
/db_xref="GI:14021761"
/translation="MQSVARSQATGPDAPVSSDSTLIYRQSRWRLTHMIWALSFF
MLLSGLQIFNARPLYIGKSGFGYNNITFEIGAVNTDAGPRGTEIFGHRFDTTGV
GWSGPRNETARGFPSWATIPSYVDLGTARVHFFPFWILTTTLVLMVASLVNNGHIR
RDALPDLDDLKRUPRIDVIDHARLKFHHTREYNTHLOKWAYGGVLFVLPMLIITGLAMS
RDNASVLPLFNEVIGGRQARTIHTFVMLLVLVGFFI IHLMLAAGFINEIRSITGM
YRTDPPAHDHDKTTTSA"
10975..11817
/gene="mlr0879"
10975..11817
/gene="mlr0879"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48373.1"
/db_xref="GI:14021762"
/translation="MTTRPRGVDMAKFEISNFOISRKRFLTAASLGASGIMLSGC
DAFDSQUSIGSGRSFLENANGUTYRAQRLAGRDALAEFTADIRQORPNGVTAP
DDDTYKGLANNFADWRLEVSGLVKPRLSLRQOLQNPSTQI TRHDCVGEWSCIAK
WTDTPLSLDQAVPQARYVMFHCIDTIDRSLGDIKYITGIDIDARHPQTIILAY
GLNGKLPVENGAFLRVRVERQLGYKMPKLYKIELVDGFANVGGGRGGYVEDNGYDM
YGI"
complement (11866..12867)
/gene="ml10880"
complement (11866..12867)
/gene="ml10880"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48374.1"
/db_xref="GI:14021763"
/translation="MGELIMSAPVAGLTSKNRIAAEDVLMRLRDVDFADGVVTRGEAEA
LPALDQATQDKCAEWAPFVEAVTDIVHOEKSGYISQDNADWLVRTISRDGWDVSR

```

```

Alignment Scores:
Pred. No.: 2,03e-57 Length: 349619
Score: 766.50 Matches: 183
Percent Similarity: 56.42% Conservative: 129
Best Local Similarity: 33.09% Mismatches: 204
Query Match: 27.59% Indels: 37
DB: 1 Gaps: 13

US-09-868-987-14 (1-552) x AP002996 (1-349619)
Qy 1 MetValSerSerProTleLeuAsnValPro---LeuLysAsnHisAlaSerValSerGly 19
Db 202374 GTGATCTCGCGCGCGAGATCCGCGTAACCGATCTCTCGCGCGCGACCGCGAGATCTCAGGC 202315
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 202314 AATTTCACCCGCCAGACGCCAATGATCTGCGCGTGTCTTTTTCGAGCGGTGCGCTCCG 202255

```

QY 40 PheValProGluValLeuSerGluClnThrIleSerSerAspLeuGlyLysLysGlnCys 59  
Db 202254 GCGAGCTGACGGTATGACAGAACGCACTCTCGGTCCGGCCCTTGGCCAGGATTCGATT 202195  
QY 60 ThrGlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerVal 79  
Db 202194 CATCGCGGCAAGGTGGCGGATCATCGTTTCGATCTCTCGTTCGTTTCATGTTTCGTC 202135  
QY 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeu 99  
Db 202134 GCCTACGGCTCTCTCGGTTCTCGCCAAACATCGCGTGGCGGTGCACGTGGCATGATC 202075  
QY 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119  
Db 202074 GTCGGCTGTTGTCTCGCTTGGGCGGACGCTGACCTTCCCGGTATTCGCGGTATCGTG 202015  
QY 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleAtrGluGlu 139  
Db 202014 CTGACCATCGGATGGCGGTGATTCACACGTGCTGATCTACGAACGCATCCGTGAAGAA 201955  
QY 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159  
Db 201954 CGACGGGTGGCGGTGATCGGCGATCGACACCGGCTTCTCGAAGGCGCTGGCG 201895  
QY 160 AlaIlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAsp 179  
Db 201894 ACCATGTCGATTCACAGTTACGTGCTGCTGATCCACCGGCGGTGCTGTCTATCTCGG 201835  
QY 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPhe 199  
Db 201834 ACCGGCCGGTGAAGGGCTTGGCCATCACCTACGCCATCGTATCTCGACACGGTGTTC 201775  
QY 200 ThrAlaLeuPheMetThrLysPhePhePheMetLeuTrpMetAsnLysThrGlnHisThr 219  
Db 201774 ACCGCTTTCACCTTACCGCGCTGCTGTGTCGATCTGGTTCGCGCGCGCTCCGAAG 201715  
QY 220 GlnLeuHisMetMetAsn---LysPheVal-----GlyIleLysHisAspPheLeuArg 236  
Db 201714 GAATTCGCAAGGCGCGGTGACCTTCATCCGCGGCGACCAAGATCCCTTCATG--- 201658  
QY 237 GlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeu 256  
Db 201657 GGTATCCGCGC---TGAAGCTTCGCGTGTGCGATTCCTGTCGGTGTCTCGTGGTG 201601  
QY 257 GlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThr 276  
Db 201600 CTGTTTCATGACGTCGACATCAATTACGGCATCTCACTCAAGGCGGCTCGATGTCGAG 201541  
QY 277 PheAsnProLysGluHisGlyLeuSerAspValAlaGlnMetArgGlyLysValValHis 296  
Db 201540 GTGAAGTCCAAGAC---GGCAGCGCAATCTCGCGATATCCGCGGCGGCTGATGGAG 201484  
QY 297 LysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSer 316  
Db 201483 -----CTCAACATCGCGCAAGTGCAGGTGCAGAGTTCGCGCGCGC 201442  
QY 317 GluIysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAla 336  
Db 201441 -----AACGACGTGTCGTCGCGGTGCGGCGGCGGATGGCGG 201403  
QY 337 SerLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArg 356  
Db 201402 GAGAAATCCGAGACGCGTATCGAC-----AAGTACCGCGGCGGAG 201361  
QY 357 ProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSer 376  
Db 201360 TTGCAGGACCAAGTAT---GATTTCGCGCGGTTCGAGTCTGGGCGGCGGTTTCGCG 201304  
QY 377 LysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle 396  
Db 201303 GAATGCGCAAG-----CAAGGCACCATCGCCATGCTGTGTCGCGCTGTTGGC 201256

QY 397 IleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrCalapheSerAlaValCysAla 416  
Db 201255 ATCTCGTCTATGCTCGTTCGTTTCGAGTGGCAGTTTGGGTCCGCCCATCTGTTGGC 201196  
QY 417 LeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLys 436  
Db 201195 ACAGTGCAGCATGTGTCATGACG-----CTCGGTTTCTTCGTCATCACC 201151  
QY 437 LysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeu 456  
Db 201150 GGCTCTCGAATTCACACAGTCTGCTGGCGGCGATCTGACCATCATCGGCTATTCGCTG 201091  
QY 457 AsnAsnThrLeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThr 476  
Db 201090 AACGACACATCGTGTCTATGACCGCTCCGAGAGGATCTCGAAAATACAGCGAATG 201031  
QY 477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496  
Db 201030 CCGCTCCGCGAGCTGTTGAACAACGCCATCAACGAGACGCTGTGCGGAACGAGCTGACC 200971  
QY 497 ThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyClySerSerValPhe 516  
Db 200970 TCGTGACACAGATCTGCGTGTCTGCGCGCTGTCTGTCGCGGAAAGTCAATCCCG 200911  
QY 517 AsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIle 536  
Db 200910 TCCTTCACCATGCAATGCTGTTCGCGTGTCTGCTACCTATCTTCGATCTTCATC 200851  
QY 537 AlaProProLeuLeu---LeuPheMetValArgLysGlu 548  
Db 200850 GCGCGCACCTGCTGATCTCTGTTCAAGCTCGCGCGCAG 200812  
RESULT 11  
AP005952/c 300600 bp DNA linear BCT 28-MAR-2003  
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 18/31.  
DEFINITION AP005952  
ACCESSION AP005952 BAO000040  
VERSION AP005952.1 GI:27352868  
KEYWORDS Bradyrhizobium japonicum USDA 110  
SOURCE Bradyrhizobium japonicum USDA 110  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobium.  
REFERENCE 1  
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110  
JOURNAL DNA Res. 9 (6), 189-197 (2002)  
MEDLINE 22484998  
PUBMED 12597275  
REFERENCE 2  
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110 (supplement)  
JOURNAL DNA Res. 9 (6), 225-256 (2002)  
MEDLINE 22485002  
PUBMED 12597279  
REFERENCE 3 (bases 1 to 300600)  
AUTHORS Kaneko,T.  
TITLE Direct Submission  
JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; 2-6-7  
Kazusa-kamata-ri, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/rhizobase//,  
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

FEATURES  
source

Location/Qualifiers  
1. 300600  
/organism="Bradyrhizobium japonicum USDA 110"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:224911"  
/strain="USDA110"  
/db\_xref="taxon:224911"  
/complement (27. . 917)  
/gene="bll14605"  
complement (27. . 917)  
/gene="bll14605"  
/note="ORF\_ID:bll14605  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49870.1"  
/db\_xref="GI:27352869"  
/translation="MFDECTMSKTTTIAILAPGAMGSVARRLSENGARVLTSLK  
GRSSTAKRAADAGMGADDEAIAEADIILSIIPPGEAVALAEALALIVRRAKKPV  
VDCNAVNDTVLRIEIEIGSAQAPFVDAGIIGFPQGGKSPAFYMSGEALRGDVAVLK  
ELGDMRIVEGPGVGAALKMSYAGIVKGLAGIGSAMVIAATKAGAADALRDELALSO  
PAVLARLEVALPDMIPKAYRWVAEMREISGFLGADHPASQIYEGFARWFEHLAEDANG  
EAVDAGLMKFAASIAQOKA"  
complement (1002. . 1772)  
/gene="bll14606"  
/complement (1002. . 1772)  
/gene="bll14606"  
/note="ORF\_ID:bll14606  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49871.1"  
/db\_xref="GI:27352870"  
/translation="WTRDDAPALDPIPEHTGLMGRIRNYFLTGLVVTGPIAITLY  
LVWFVWDSVVRFPFLAYRPETLYPVVPGGLIVAFPTLLVGLFLANLIGRL  
VDVGETFUGRIAPVARIYGLKQVFTELFGSGSSFRKVGLEFPSPGMSIVLISOS  
PNEEVARSLPQGEHVSFLPCSPNPTTGFYFPKSKIIVEVDNMSTEDATLINSAGV  
VQPGAADPKKAALAGNAARANANASALQOPAKVE"  
complement (1852. . 3678)  
/gene="glms"  
/complement (1852. . 3678)  
/gene="glms"  
/note="ORF\_ID:bll14607"  
/codon\_start=1  
/transl\_table=11  
/product="glucosamine-fructose-6-phosphate  
aminotransferase"  
/protein\_id="BAC49872.1"  
/db\_xref="GI:27352871"  
/translation="MCGIVGILGREPVAEQLVDSLKRLEYRGYDSAGVATLEGHRLR  
RRAEGLKNLEKLEAEPLKGTGTGIGTRWATHGKPTVNNAPHATERVAVWHNGIIE  
NRELRELEKNGTVFHTTETDETEIVLHLVDDLTRGNKPVFAVKLALARLGAALGP  
IFAGDNDLMICARNGPPLAIGYDGENYLGSDAIALGPTDTIISYLEDGDWVLTRNS  
ATIFDKGHAOREKIRHAASLTVDKANYRHFMAKEIHEQPEVVGHGTLARYVDYATE  
RVSLFVLFPDFKNIQIRNITACGTASYAGFAVYKWFEPARLFVEVDVASEFVREA  
PLRGDLAIFISQSGETADTLAALRYAKAGVHTVAVNVPTSTIARESVTLPTLAG  
PEIGVASTKATCOLMLVANLAIAGKARGELSDDETKLVHGLVEIPLMSDALTTE  
LQIEKLAREIAKSDVLVLRGTSFPPLALEGALKKEISYTHAEGYAAGELKHGPIAL  
LIDETPVVVIAPYRPFVKTYSNQEVNARGGKILMTDAKAGEATEVSLVTIWMFD  
MAAATPMVYAVPVQLLAYHTAVTNGTDVDQPRNLAKSVTVE"  
complement (4121. . 5476)  
/gene="glmu"  
complement (4121. . 5476)  
/gene="glmu"  
/note="ORF\_ID:bll14608"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylglucosamine pyrophosphorylase"  
/protein\_id="BAC49873.1"  
/db\_xref="GI:27352872"  
/translation="MTAKRSLLTIVLAGEGTRMRSHLPKVLHPVAHOTLLAHVLTAAAP  
KGTGTSVAVTGPQVADSEARIRPDVTFVQAERLGTAAHVLAALAEATARGVDDL  
LIAFGDTPLIASFTARLAPLAAALGAFRAADPAGYGRFIVEGDRILVAIREQA  
DASADERKIDLCNAGVMAIDGRRALAILDKIGNANSKGEYVLTDAVEIVRQGVESVV

gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS

IETSEDEVRGINTKAOLAEAESVMQARLRKAAMEAGVTLLIAPETVYLSADTVFGKQVT  
IEPFWIGPGVSIADGTVIHFSFHI VETTLGRNVISGPYARLPGTSLGQDGAIGNFV  
ETKAAITLGAQKUNHLSYIGDATVGANSNIGAGTICTNYDGFKKHKTIIIGOGAPVGTN  
SSLVAPVKINGAYIGSGSVITRDVPDDAMALERNOQTIREGGNARYENKTRGKKFE  
K"  
5649. . 7028  
/gene="bll14609"  
5649. . 7028  
/gene="bll14609"  
/note="ORF\_ID:bll14609  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49874.1"  
/db\_xref="GI:27352873"  
/translation="MAKDSPLADAGTGTGGLFSGLLAEESFDRMMWRLLSGWGV  
AAGVAVVYMANQALQWRDDQNASADLVQRADRLQLLTRESONEARRLAAAVETLN  
AADRLYSRTVLEQGLDSVTGAIRQSAKQPDTAPDAQPAAPSVAPASMPSPSSD  
KPRAAAKEQAEKATKDTVKETAKESPSPPOSAASALVQOSPATTALPMLPLVPS  
KSIWAPDPAASKLTQPTETKETAETAEKKPEPTAPTEITAAAKPPPEATESEAPA  
IAYQOTRFAIDLGANSIDGLRALMRGVTKSNPEVATLRPIIMIKEGTTGIGMOLRLG  
AGLLINAAAALCAGLAENDRHCEITVFDGQRLSMRGSGSEKGEKSDQVQEKNAEK  
NPDAAPQOETAPAPAPTAKPKRRRSYSSKRSKREBPAPAPAPAPAKPETAAAG  
STLSSPFR"  
complement (7106. . 7345)  
/gene="bsl14610"  
complement (7106. . 7345)  
/gene="bsl14610"  
/note="ORF\_ID:bsl14610  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49875.1"  
/db\_xref="GI:27352874"  
/translation="MADGGAEVDVEQVQDRKSFVEFVAAMRNELLVGAKGMENVDLAD  
FLSAMTAAGDHSHPANPNPWRHAAALMWAKLYE"  
7466. . 7846  
/gene="bll14611"  
7466. . 7846  
/gene="bll14611"  
/note="ORF\_ID:bll14611  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49876.1"  
/db\_xref="GI:27352875"  
/translation="MTKKPRLTPYQVQLIVGVFVAVGYAIYLRYLGIENSNVGLAC  
DAGLQTLMKCARSVSTVLFKNSVFGIVALVATLNLMRPSIVLLTVGVVAAAGLGIVLY  
NVVLGSLAIGLLILGFARPAPATA"  
complement (7806. . 9470)  
/gene="ndvC"  
complement (7806. . 9470)  
/note="ORF\_ID:bll14612  
putative beta (1-6) glucans synthase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC49877.1"  
/db\_xref="GI:27352876"  
/translation="MAREQVIDGASRRTEPISLRTLPLALLLSIGATAAVNMWLATPI  
TLARADIPADKQVCVYAPFGEQSPNLAWHIEADQIEODLRQLKEITDCVRYTSM  
ENGLDOVPAIAAKVGGKVLQGIWLSNRAKNFEQVALAVRLTKEYPDITTSVTVGNE  
VLRGENTADLVSFIRAVKQVVRVPTADVNEFWLRNREVTEADVFTIHLIPYME  
DVPFRKAAHAEVQIERMAVAPPGKEIILIGETWSPSEGRMGALPSTRNQARVVS  
EILSLAKAQKFRVNLIESYDQPKWRKLEGTAGGYWGLYDSVRRNLKYTPGGEPTSNFPY  
WKYMGAMGMLCALVFVAVITLRRRPTFRFTSLGVAISATTAGVLLGVGADKVVY  
EYSGVGMHLHWGALLAAGILSPIFCAQALVTRSLPTFLDLGLGSEGRKWSKLTAVLG  
LTVAVTVAIAQOTALAPVDFPRYHFTWAVPFAILMLNRPQIGVYRPIAEAVF  
ACVLASAAVYLFNEGRDNWQSLWCAMYLFLFALTMLRAREQSQE"  
9694. . 10290  
/gene="bll14613"  
9694. . 10290  
/gene="bll14613"

```

/note="ORF_ID:bl4613
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49878.1"
/db_xref="GI:27352877"
/translation="MRORVRSRNVAVLRQFAATLAVSLIALAGLAGAQAQSGTPAPD
OGKAAQPADAAADKAAQNRRTDEFAAEQVINGPAGNPECVWLGRRVLRVLRWDRDL
DTAFRLDLYDRFCPGGHIQAAFRCLTRFGGQIDPKVATLDSRHACWINPASQPK
QAAAAQAPAAPTSGNAQAQPPAASPTAPQK"
gene
10509..13178
/genes="ndvB"
CDS
10509..13178
/genes="ndvB"
/note="ORF_ID:bl4614"

```

## Alignment Scores:

```

Pred. No.: 1,85e-57 Length: 300600
Score: 766.00 Matches: 189
Percent Similarity: 53.40% Conservative: 125
Best Local Similarity: 32.14% Mismatches: 201
Query Match: 27.57% Indels: 73
DB: 1 Gaps: 12

```

US-09-868-987-14 (1-552) x AP005952 (1-300600)

```

Qy 1 MetValSerSerProfileLeuAsnValProLeuLys---AsnHisAlaSerValSerGly 19
Db 143619 GTGATCTCGCGCTGTCTCCGAGCGGATACCGCGGGTCAGGCGCGATCTCCGC 143560
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 143559 AATTTCACCGTGAATCGGCCAACGATCTCGGATCTCTCGCGCGCGCGCTGCCG 143500
Qy 40 PheValProGluValLeuSerGluGluThrLysSerSerAspLeuGlyLysLysGlnCys 59
Db 143499 GCGCGCTCACCGTGGTGGAGGCGGACCGCTCGGTCGGGCTTGGCCAGGACTCGATC 143440
Qy 60 ThrGlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerVal 79
Db 143439 GAGAGGCGAGCTTGGCTTACGTCGGCTCGATCTCGTTCATCTTCATCTCTCGTG 143380
Qy 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeu 99
Db 143379 ACCTATCGGTGTTCGGGTGTTCGCCAACATCGCGTGGCCATCAACGTCGCCATGATC 143320
Qy 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLysSerGlyLeuAlaGlyIleVal 119
Db 143319 TTGGCTGTGTCTGCTCAACGCGCACCGCTGACGCTCGCGGATCGCGCGCATCGTG 143260
Qy 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGlu 139
Db 143259 CTCACGTCGGCATCGCGGTGGATCCAACTGCTGATCTACGAGCGCATCCCGGAGGAG 143200
Qy 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159
Db 143199 CTGCGCGCGCGCGTAAACGCGATCTCGCGCATCGACCGCGCTTCAAGCGGCGCTCGCG 143140
Qy 160 AlaIlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPheLeuAsp 179
Db 143139 ACCATCTCGATTCCAACTACACACCTTCATTGCGCGCGCGGTGTGTCTACATCGGC 143080
Qy 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPhe 199
Db 143079 ACCGCGCGGTGCGGCTTTCGGGTGACGCTCGTATCGGCATCATCACCGGTTTC 143020
Qy 200 ThrAlaLeuPheMetThrLysPhePhePheMetLeuTrpMetAsn----- 214
Db 143019 ACCGCTTACCCCTGACCCGCTGATCTGTCGCGGCTGGTGGTGGAGCGCGCGCAG 142960
Qy 215 -----LysThrGln-HisThrGlnLeuHisMetMetAsnLysPh 227
Db 142959 AGCGTCCGATCTAGGAGCCGTGATGTGGAATCACACCGTTCATCGGCTCGGCTCCT 142900

```

```

Qy 227 eValGlyIle----- 230
Db 142899 CATTACTATTCTCACCGTGTTCGGCGCCCTCGCGCTTGTCTGCCCTGTTGGCGCATCGTCCC 142840
Qy 231 -----LysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySe 247
Db 142839 GGACAACACGCAATTCGACTTCACGCGATTCCCGCCGATCAGCTTCCCGATCTCCGGCGC 142780
Qy 247 rValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMe 267
Db 142779 GCTGTTCGATCGTCCCATCACGCTTCTTACCCACGCGCTGAAC-----CTCGGTAT 142726
Qy 267 tAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVa 287
Db 142725 CGACTTCAGGCGCGGACGCTTGTCTGAGGTGCGGGCCAGTCC---GGCGCCCGCGACAT 142669
Qy 287 lAlaGlnMetArgGlyLysValValHisLysLysLeuGlnGluAlaGlyLeuSerSerArgAs 307
Db 142668 CGCGCGCATGCGCACGCGCTG-----GATGGCTTGGCTCGGCGCA 142627
Qy 307 pPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAl 327
Db 142626 GTCCAGCTCCAGCAGTTCGCGCGGCGCGAGAACGTCCTGATCCGCTCGCGGAGCAGCC 142567
Qy 327 aLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpAr 347
Db 142566 GGGCGCGGATACCGCCACGAGGCGCGTG----- 142536
Qy 347 gTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSer----- 364
Db 142535 -----CAGAAGTTCGCTACCGCGCTCGCGGATTCATCCAGTA 142498
Qy 365 -LysArgAsnAlaLysPheTrpSerLysValSerLysLysLeuSerLysLysMetArgTy 384
Db 142497 CCGCGCGCTCGAGGTCTCGTCCGCGCTCGCGCGCGAGCTC-----TTGCGCTA 142447
Qy 384 rGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuLeuTyrValSerLeuAr 404
Db 142446 C---GGCATGCTCGGCTGATCTCGGATCTCGGATCTGATCTATCTCTGTGTTCG 142390
Qy 404 gPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaTh 424
Db 142389 CTTCAATGGCAGTTCGCGCTCGCGCCATGATCGCCTACGACGTCACGACATC----- 142338
Qy 424 rCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeuGlnAl 444
Db 142337 -----GTGCTGACGATCGGCTTCATGTCGATCATGTCAGTTCGATTCGACCTGACCAG 142285
Qy 444 aIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAs 464
Db 142284 CATCGCGCGCTTCGACCATCTCTGGCTATTTCGCTTACAGACGCGTCTGATCTACCA 142235
Qy 464 pArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAs 484
Db 142224 CGGTATCGTGAATGCTGCGCGCTACAGAAGATGCGATGCGCGCAGCTCTCAACCA 142165
Qy 484 pAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerVal 504
Db 142164 GTCCATCAACTCGACGCTGTCGCGCTCATCATCACCGACGTCAGCGGTGACGCTGCGCT 142105
Qy 504 uLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAlaPheIleMetThrIle 524
Db 142104 GTCGCGCTGCTGCTGTCGCGGCTCAGCCATCCACAGCTTCACCGCGGTGATGTAT 142045
Qy 524 eGly---lLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPh 543
Db 142044 CGGCGTGTGTGTGTCGCGCACCTACACCTCGATCTTCATCGCGCGCGCATCTGATCTA 141985
Qy 543 eMetValArgLysGluAsnArg 550
Db 141984 TCTCGCGCTCGCGGAGCATCGC 141963

```



/gene="iscA"  
 /locus\_tag="PA3812"  
 /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable iron-binding protein IscA"  
 /protein\_id="AAG07199.1"  
 /db\_xref="GI:9949986"  
 /translation="MWSMTAAAKHVSLEGRKGEGIRLGRVTRTCGSLAYVLFVDELAEDLVFESHGVKVIIDPKSLVLDGTDELDTREGLNEGFNNPNVRGCGCGESFNV"

## gene

complement (3507..3893)  
 /gene="iscU"  
 /locus\_tag="PA3813"  
 /complement (3507..3893)  
 /gene="iscU"

## CDS

/locus\_tag="PA3813"  
 /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable iron-binding protein IscU"  
 /protein\_id="AAG07200.1"  
 /db\_xref="GI:9949987"  
 /translation="MAYSEKVIDHYENPRNVGKLDADPNVGTGMVGAPACGDVWRILQIKNEQGVIEDAKFTYCGGSAIASSLATEMMKGLDEATIKNTTIABELALPPVKHGSVLAEDAIAKAAVDYRKOKGGL"  
 /complement (3929..5143)  
 /gene="iscS"  
 /locus\_tag="PA3814"  
 /complement (3929..5143)  
 /gene="iscS"

## gene

complement (3929..5143)  
 /gene="iscS"  
 /locus\_tag="PA3814"

## CDS

/note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Alt protein name: iscS protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="L-cysteine desulfurase (pyridoxal phosphate-dependent)"  
 /protein\_id="AAG07201.1"  
 /db\_xref="GI:9949988"  
 /translation="MKLPIYLDYSATTVPDPRVAQKMSCELLMDGNFGNPSRSHVFGWKAEAVENARQVAELVNADPREIVMTSGATESDNIAIKGVAHFYSGKKGKHIITSKI EHKAVLDTQRLREGEVTVLEPEDGLITPALVEAALRDDTLVSMVHWNIEIGTVNDIAAIGELTRSRVLPHVDAQAOSTGKVEIDLKLDVLMFSFAHKTYGPKGIGALYVRRKPRVRIEGMHGGHGERMSCTLAHQIVGMGEAFRIAKSEMAQENARVIALDRPFPAQIDGLEELYINGMTSRVPHNLNVSNFVYGESLIMALKOLAVSSGSGASLSLEPSYVLRALGRNDELHAHSSIRFTFRFTTEBEEIDYAAKKVVEAVSKURELSPLWDMYKEGVDSLQVWQAH"  
 /complement (5173..5664)

## gene

## Alignment Scores:

Pred. No.: 7,17e-59 Length: 14652  
 Score: 762.50 Matches: 182  
 Percent Similarity: 55.42% Conservative: 130  
 Best Local Similarity: 32.33% Mismatches: 183  
 Query Match: 27.45% Indels: 68  
 DB: 1 Gaps: 13

US-09-868-987-14 (1-552) x AE004799 (1-14652)

Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
 Db ATCATCAGCTGGCGACCATCCAGTCGCGCTGGGCAACAGTTCCGATCACCAGGCTC 10705  
 Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLysSerGlyAlaMetSerPhe 40

Db 10704 GATGGCCCGCGGAGTCTCTCCGAGCTGGCCCTGCTGCTGCGCGCGGCTGGCTCGC 10645  
 Qy 41 ValProLuvValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
 Db 10644 CCGATGTACTTCCGCGAAGAACGACCATCGGCCGAGCTGGGGCGGCGAACATCGCC 10585  
 Qy 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
 Db 10584 AAGGTATCGATCGCTCGCTCTGGGCGATGCTCTTGTATCGCTGTTCATCATCGTCATC 10525  
 Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr 100  
 Db 10524 TACCGTTTCTTCGCGGTGATCGCCACCGCTTGGCTTCAACATGGTTCATCGTGTG 10465  
 Qy 101 AlaAlaLeuGlnTyrIleuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 10464 GCGCTGATGTCGATCTCTGGCGGACCCCTGACCTCGCGCATCGCGGTATCGTGTG 10405  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleAtrGluGluPhe 140  
 Db 10404 ACCATGGCATGGCGGTGGAGCCACGCTGCTGATCTTCTCGCGGATACCGGAGGACTG 10345  
 Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 10344 GCCAATGGCATGTGCGTCCAGAGGGGATCCACGAAGGCTTCAACGCTTCCACGCG 10285  
 Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
 Db 10284 ATCTCTGACGCCAACCTGACCTCGCTGCTGGCGGATCTCTTACGCCATGGGCGACC 10225  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 10224 GGCCCGGTGAAGGATTTCGCGTGACCATGTCGCTCGGATCATCACCTCGATGTTTACG 10165  
 Qy 201 AlaLeuPheMetThrLysPhe-----PhePheMet 210  
 Db 10164 GCCATGCTGTCACCCGCGCAATGGTCAACCTGATCTTCGGCGGCGTGTGACTTCAAGAAG 10105  
 Qy 211 LeuTyrMet-AsnLysThrGln---HisThrGlnLeuHisMetMetAsnLysPheValG 229  
 Db 10104 CTGTGATCTTAAGGGGACTGAGATGAATATCAAGATCGGTACCATCAAC---TTTATGGG 10048  
 Qy 229 YlleLysHisAspPheLeuArgGlyCysLysLysLeuTyrAlaValSerGlySerValPh 249  
 Db 10047 CATCCGCAAC-----GTGCGGTTTCCGCTCAC 10021  
 Qy 249 eLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTyrPheSerVal-----LeuG 266  
 Db 10020 CTTGATC---CTGACCGTCTCGCCCTGGCGAGCTGTTTCAACAGGGCATCAACTTCGG 9964  
 Qy 266 yMetAspPheLysGlyGlyTyrAlaPheThrPheAsn?oLysGluHisGlyLysSerAs 286  
 Db 9963 CTTGACTTCAACCGCGGTACGCTGATCGAGTGTACCTACGAACAACCG-----GCCGA 9910  
 Qy 286 pValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerAr 306  
 Db 9909 TCTCGCGAAGTGGCGTGGTCACTGGTC-----GGCGCCGCTATGA 9868  
 Qy 306 gAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLy 326  
 Db 9867 GGACGCGGTGTCAGAGCTTCGGCGACCGCGCGAGCTGCTGCTGCGGATGCCAGCGA 9808  
 Qy 326 sAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTr 346  
 Db 9807 GGACCGGAACCTGGGCAAGAAGTCTGCTACCGCTTGCAGCGGCGCAT-----9759  
 Qy 346 pArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysAr 366  
 Db 9758 -----GCCGCGCAACCCGCG 9745  
 Qy 366 gAsnAlaLysPheThrSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAl 386  
 Db 9744 GAACCTGAAGCGGTGTCAGTACGTGCGCCCGAGGTGGTGAAGAATACTGCGCGACCGG 9685

Qy	386	aThrileGlyLeuLeuGlyAlaLeuAlaIleileLeuLeuTyValSerLeuArlpheG1	406
Ds	9684	CGGCTCGGCATGCTCTCGGCGTGGCGGCATCTGCTGACGTGGCTTCCA	9625
Qy	406	uTrpGlnTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAl	426
Ds	9624	GTGGAAATTCGCGCTCGGTCGATCTCTCGTGTGACGATCGCATCGTATGGG	9565
Qy	426	aValLeuPheIleAlaHisPhePheLeuLysIleGlnIleAspLeuGlnAlaIleG1	446
Ds	9564	CGTGCTG-----TCGTTCTTC-----CAGGTGACCTTCGACCTGACCGTCTCTCGC	9520
Qy	446	yAlaLeuMetThrValLeuGlyTySerLeuAsnAsnThrLeuIleIlePheAspArg1l	466
Ds	9519	CGCTGCTGGCGGTGGTCTGCTGCTGAACGACACCATCGTGATCTTCGACCGGT	9460
Qy	466	eArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHisValLeuVa	482
Ds	9459	GGCGAGAACTTCGCGCTGCTCGCAAGGCCGATCTGGTCGAGAACCTGAACATC	9405
Qy	482	IAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSe	502
Ds	9404	-----TCCACACCGACCGCTGTTGCTACCATCGCCACCTCGGTTTCCACGCTCT	9352
Qy	502	rValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMe	522
Ds	9351	GGCTATCGCGCGCTGCTGTTCTTCGGCGGCACATCTGTTGCTGTTTCTCCATCGCCT	9292
Qy	522	tThrileGlyIleLeuLeuGlyThrLeuSerSerLeuTyIleAlaProProLeuLeuLe	542
Ds	9291	GTTCTGCGTGAATGGCGGTACCTACTCGTCGATCTACATCGCAACGCTGTTATTCAT	9232
Qy	542	uPheNet	544
Ds	9231	CTGGCTG	9225
RESULT 13			
AE016777			
LOCUS	AE016777	300861 bp	DNA linear BCT 12-DEC-2002
DEFINITION	Pseudomonas putida KT2440 section 4 of 21 of the complete genome.		
ACCESSION	AE016777	AE015451	
VERSION	AE016777.1	GI:26557020	
KEYWORDS	Pseudomonas putida KT2440		
SOURCE	Pseudomonas putida KT2440		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
REFERENCE	1 (bases 1 to 300861)		
AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Uterback,T., Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J., Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoft,A., Tummier,B. and Fraser,C.		
TITLE	Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440		
JOURNAL	Environ. Microbiol. 4 (12), 799-808 (2002)		
REFERENCE	2 (bases 1 to 300861)		
AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Uterback,T., Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J., Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoft,A., Tummier,B. and Fraser,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		

source	1. 300861	/organism="Pseudomonas putida KT2440"
		/mol_type="genomic DNA"
		/strain="KT2440"
gene	114. 2351	/db_xref="taxon:160488"
CDS	114. 2351	/locus_tag="pp0788"
		/locus_tag="pp0788"
		/note="identified by Glimmer2; putative"
		/codon_start=1
		/transl_table=11
		/product="conserved hypothetical protein"
		/protein_id="AAN66413.1"
		/db_xref="GI:24982192"
		/translation="MATSINRHSFDDFSALSCKTPCKESRMOKEGVVPLAATIDRGG
		DLRHPCLUPVLQVRDKAQLQKQLOLFDNADDTLUFEMADKAFDGGDQNLVPEAMR
		RLKRSIERGFLDTFDLAFRAGQVLDLLANLAQPGKPSKAQMERAAAEIYGVAVR
		LSRDGVALQQLGLRLQALLDRPLYEQHNPLGPAALCRYFLDAGNLGVRLKVLVK
		LFERYVLDVDVITYGEANQLAAAGVLPQLQAPRRRAEDRRMRVRPGPANLGOEBA
		DAGQALFASLQALLAPVRCGFAPRLQAAQAQPISTADLLRLLSHLQRYVPATYEA
		DFELGQLEQLLRLSVRSRRRLHVAEDMNLVGLLFAFIHNDNLPLASRLALIT
		RLHPLIKVALLDKGLFSRSSHPARRLLNEIAGAAIGWEGSECCORSLHMRVERIVQ
		RLNDFADDAFLFAELDDLELAFNQDERRNELLEQRTDAEGRARMAQARQOVHV
		LNORLRGVLQVQVQMLVQASQVLLAWLKQGEASQAWRDVQTMMLIASITPPH
		EPQALLQVPLGLKALRDGLASVALDAVATREFFLQELQHLRACACKEALPSGEGS
		LGVLVQEELVLTIAEPPACVPLHTADQAAALRQARLRIGTHVVELDEDEPLRCKL
		VARIDSSDLRVFANRTGMKRVWSSASLAQALQORHEVRVLDRLFLERALEAVLESUR
		QGQAQ"
	2440. 3012	/gene="ampD"
		/locus_tag="pp0789"
	2440. 3012	/locus_tag="pp0789"
		/gene="ampD"
		/locus_tag="pp0789"
		/note="similar to GP:4079828, GB:K02765, SP:P01024, PID:179665, and PID:181130; identified by sequence similarity; putative"
		/codon_start=1
		/transl_table=11
		/product="N-acetyl-L-anhydromuramyl-L-alanine amidase AmpD"
		/protein_id="AAN66414.1"
		/db_xref="GI:24982194"
		/translation="MQLDYATGFWGHGITHCPSNPNARPEGESISLLVTHNISLPP
		ACFGTKVQOFFQNRLLAPDEHPYFASINHLTVSAHLVEVDGAVTQVSLHDRAWHAG
		VSSFDRGKCNDFSIGIEGTDDLPYDAQYAVLEQUTRHSAPFAIDNRIQGH
		DIAPQRKTDGPAFDMRSRYRQALQHNEDKA"
	3009. 3839	/locus_tag="pp0790"
	3009. 3839	/locus_tag="pp0790"
		/codon_start=1
		/transl_table=11
		/product="inner membrane protein AmpE"
		/protein_id="AAN66415.1"
		/db_xref="GI:24982195"
		/translation="MSFLVLLALWKEKFSALRHQVORDGFFIGELVRLERNKGVHPW
		WTUAILVAPVALLVLLHLEVPAYGLLAPVHLLVLIYSLGRGDNAKSLGAPRDAA
		RRGDDQAALHAAERDLGVADEPHSLVLRVQGNLLQVQYGFYFVYVLPFGAAL
		AYRLLALCSEHSKQPALKARAEQRLHMDLVPVRLALSALVGNFLAVTRVLMHLEVL
		NWHSAGHLVARVGRIADTIPEEDSQRLGLRLDSLWELLRLCAVLWVAGFALWTVLV
		"
		/complement(3884..4660)
		/locus_tag="pp0791"
		/complement(3884..4660)
		/locus_tag="pp0791"
		/note="identified by match to TIGR protein family HMM TIGR00010"
		/codon_start=1
		/transl_table=11
		/product="deoxyribonuclease, TatD family"
		/protein_id="AAN66416.1"
		/db_xref="GI:24982196"



/translation="MRLDTHLDFPDDADRRLLANAARGLRMVVLGVYQANF  
QRVWELACAEPLFAELGLHPVYLDOHPEHLVOLRELRLNPRLCVAGRGGLDY  
YLALDTRQQLAEQAQACDFELPALLVRRSHQAVIATLKRYKPARAGVTHAFAP  
GSYEAREYIKGLFGLGGAGTWPQALURKRLIARPLESVILETSDPMPAPAMYGP  
VRNPEHLPEIATALAHVMGUDVEVLAQASTRNACTLFGW"  
complement (4663. .5658)

gene

/gene="frub"  
/locus\_tag="PP0792"  
complement (4663. .5658)

CDS

/gene="frub"  
/locus\_tag="PP0792"  
/note="similar to GP:9949716; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="fructose transport system repressor Frur"  
/protein\_id="AAN66417.1"  
/db\_xref="GI:24982197"

/translation="MKLSDIARLAGSVTTASVYVINGKAEORISNSTVRRVAVFA  
HGFTPNQAGLSRHRRTLGLFILPDLENPSYARIKQEGGARQGYQLLIASSDDQ  
PDSERQQLFRARCDALFVASCPEPDDSYRELQDKGUPVIAIDRRLPDAPFCSVI  
SDSDASRQLAASLLSAPRSIALIGARPELSVSQARAGGFDEALOGYTGEVRYOGE  
AFDRRCQRLMQQLIDGLGPDALVTTSYVLLQGVFDTLQARVPDSRQLQGLTFGDN  
QLLDFLPLPVNAMAQQHGOIAATALELALAAIEKRYEPGVHAGVTRFKQISVA"  
5943. .8795

gene

/gene="frub"  
/locus\_tag="PP0793"  
5943. .8795

CDS

/gene="frub"  
/locus\_tag="PP0793"  
/note="similar to SP:P23388; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphotransferase system, fructose-specific  
EI/HP/ETIA components"  
/protein\_id="AAN66418.1"  
/db\_xref="GI:24982198"

/translation="MLELANEQIANGQKADKABRLALLDLRLVADGLVAGYLQGLQ  
AREAGSTFLGQGAIPHGTPQTRDLVYATGRLLQFPFGVMDGQGVYLAIGIAR  
SDEHLRLQLTRALGETDLAEALRRASSAEALLKLQGAPOALALDAQLVGLNPAE  
DFDLAWRGALRLQRCADVSGFAALVQQAEPPLBGLGMLHSEROVRQPGLAFTIP  
QVPLRYRDPONGPLFCLASCAHEALLERLCEVLI EGRGOVLYQATSSRAVLEVLGG  
EQRADMPASRVLANPHGLHARPAKLAQLAKGEGEIRVRLVDSQAPVSVKSLKL  
LSICARGQALVELVAESFIADALPVLIAATEQGLGEEVEPLOSVAPIADDPVFLQ  
APAGSRIGQGAAPATAGSHAHCVEREFDYPLRGESCALERKQLEALATVNGELQ  
ALVLRSDKAIGEIVTQHEMLADPALTDDVEQLAQGESAAAMWAVIEAARQCEAL  
HDALLAERADLRIGRRLVLAQLCCVQAQVEPQYVYLVMTVEPVSARLDPNRVAG  
IVTAQGGATAHSATVARALGIPAVVGAASITLLESCTPLLLDQGRGVSVAPPADEL  
ORALAEERDLREORLOAWANFEPAITRDGHAVEVFANIGDSNGIAKVYQGAEGVL  
LRTLEITFMHPQAPDQVQAEYRRVLDGLDRLPVLVTLVDGDKPLPYWPIAAEEN  
PFLGVGRVRLTLQRPQVMEQDLRALRAADQRPRIIMPVQGVHWEAREARVERLR  
AEIPEVADQLGIMVEPVSAAALAAQAREVDFFSITGNDLTQYTLATDRGHPSLSQA  
DGLHPVLSLIDMTVRAAHAGKHWGVGCLAADPQAVAVLLGLDVLDELVSARSIAE  
VKALVROADHOTARALAREALQODSAAAVRALVERY"  
8796. .9743

gene

/gene="fruk"  
/locus\_tag="PP0794"  
8796. .9743

CDS

/gene="fruk"  
/locus\_tag="PP0794"  
/note="similar to GB:L20852, and PID:306772; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1-phosphofructokinase"  
/protein\_id="AAN66419.1"  
/db\_xref="GI:24982199"

/translation="MAKILTLTLNLPALDITIGLTLRPGQVNSQAHSHAGKGLNV  
AQVADLGHVSVVGFGGDNLOQFEALIDRGTDFCFVRPGETSRNLIKLEADRGV  
TDINGQGVDEVAARSALLRLQIAFCHDAVVVAGSLPRGISADFWQLLERLKAQV  
LKVYDSSGALRYLQSPWLPKNTPEELGVLGAVDNLTKQRAAAKRLSDSGVEH  
VVSAGSQGVSWFSDRLAQARPPKRVAVTGVAGDSLVAGMVHGLLLAEAPQTLTR

ATAIAQAQVTVQGVGFIHREHLAQLAEAAVQLTEQBGCR"

9740. .11482

gene

/gene="frua"  
/locus\_tag="PP0795"

CDS

9740. .11482  
/gene="frua"

/locus\_tag="PP0795"

/note="similar to GB:J04213, GB:L34219, SP:P12271,  
PID:L90854, and PID:598229; identified by sequence  
similarity; putative"

/codon\_start=1

/transl\_table=11

/product="phosphotransferase system, fructose-specific  
IIBC component"

/protein\_id="AAN66420.1"

/db\_xref="GI:24982200"

/translation="WNIAIVTACPNQGVSVLSARLLSAAQRGMSTSEVQNTFHP  
ERQLSAQIAEADWLIVSTGPVDLRFVKGELYQSTPSQALADRGFLDEAANAEL  
LAAVTCAPAAASSAGARI VAVTACPTGVATFMAAEALQAAQOLGQVLTVEQTGSVG  
ARNPLSAEAIADADVLLAADIEVTPARFAGKRIYRCGTGIALKQARATLDKALAQAT

## Alignment Scores:

Pred. No.:	4.92e-55	Length:	300861
Score:	739.50	Matches:	184
Percent Similarity:	54.53%	Conservative:	123
Best Local Similarity:	32.68%	Mismatches:	184
Query Match:	26.62%	Indels:	72
DB:	1	Gaps:	13

US-09-868-987-14 (1-552) x AE016777 (1-300861)

Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20

Db 69350 ATCATCAGCTGCGCACCACCATCCAGTCGCCGCTGGTAG3GCAGTTCGCCATCACCGCTTG 69409

Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40

Db 69410 AATGGCCAGGGCAATCGTCCGAGCTGGCCCTGCTGCTGCGTCCGCTGGTCTGGCTGCA 69469

Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuLysGlnCysThr 60

Db 69470 CCGATGTACTTCGCTCAAGACGACCTACCATTTGCCCAA3CCTGGGTCTGCAACATCACC 69529

Qy 61 GlnGlyIleSerAlaCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 80

Db 69530 AAGGGTATCGATCGCTGCTGGGCGATGCTTTCGCTGCTTCATCATCGCCATC 69589

Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr 100

Db 69590 TACCGCGCTTCGGTGTGTGCGCCACCATCGCCCTGGCGGCAACATGCTGCTGCTGCTG 69649

Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120

Db 69650 GCCTTCATGTGCTGCTGGGGCGACCCCTGACCTGCGCGGTATTGCCGCGATCGTGTGTTG 69709

Qy 121 AlaMetClyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140

Db 69710 ACCATGGGTATGGCGGTGGACGCCAACGTCGTGATCTTCGCGTATTCCGGAAGAGATC 69769

Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysClyTyrThrLysAlaPheGlyAla 160

Db 69770 GCGGCTGGTATGTGCTGGTGAGCGGCCATCCACGAGGCTTCAATCGCCCTACTCGGCA 69829

Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180

Db 69830 ATCGTCGACGCCAACCTGACCAACCCCTGTTGGTCGCGCGCATCTCTGTTTCCCATCGGTACC 69889

Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200

Db 69890 GGCCCCGGTCAAGGCTTTCGGGTCCACATGTCCTCCGGGATTTTACCTCGATGTTTCCACC 69949

Qy 201 AlaLeuPheMetThrLysPhePhePhe-----Met 210



69950	Db	GC	CGT	CAT	GGT	GAC	CGT	TGGT	CA	ACCT	GAC	CTG	TGG	CGG	CGG	CGT	GAC	AT	CA	AG	AAG	70000		
211	Qy	Leu	Trp	Met	-----	Asn	Lys	Thr	Gln	His	Thr	Gln	Leu	His	Met	Met	Asn	Lys	Phe	Val	228			
70010	Db	CT	TGG	G	GT	T	TG	AGG	G	AG	CT	CG	G	A	T	G	A	A	C	C	A	T	70051	
229	Qy	Gly	Ile	Lys	His	Asp	Phe	Leu	Arg	Gly	Cys	-	Lys	Lys	Leu	Trp	Ala	Val	Ser	Gly	Ser	Val	248	
70052	Db	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	70078	
248	Qy	l	Phe	Leu	Leu	Gly	Cys	Val	Ala	Leu	Gly	Phe	Gly	Ala	Trp	Asn	Ser	Val	Leu	-----	G	I	266	
70079	Db	C	A	C	G	T	C	T	C	A	C	G	T	G	T	T	G	C	A	A	G	G	C	70138
266	Qy	y	Met	Asp	Phe	Lys	Gly	Gly	Tyr	Ala	Phe	Thr	Phe	Asn	Pro	Lys	Glu	His	Gly	Ile	Ser	As	286	
70139	Db	G	C	T	G	G	A	C	T	T	C	A	C	G	G	G	G	T	A	C	C	T	T	70192
286	Qy	p	Val	Ala	Gln	Met	Arg	Gly	Lys	Val	Val	His	Lys	Lys	Leu	Gln	Ala	Gly	Leu	Ser	Ser	Ar	306	
70193	Db	C	G	T	C	A	A	G	G	T	G	T	C	G	C	G	A	G	T	T	G	T	C	70235
306	Qy	g	Asp	Phe	Arg	Ile	Gln	Thr	Phe	Gly	Ser	Ser	Glu	Lys	Ile	Lys	Ile	Tyr	Phe	Ser	Asp	Lys	326	
70236	Db	-	G	A	G	G	C	G	T	G	T	C	A	C	T	G	C	T	G	C	T	T	70294	
326	Qy	s	Ala	Leu	Ser	Tyr	Thr	Lys	Ile	Arg	Ala	Ser	Leu	Leu	Lys	Leu	Thr	Ile	Met	Ser	Tr	346		
70295	Db	C	G	A	T	C	C	C	A	A	G	T	A	G	C	C	T	G	C	A	A	G	70343	
346	Qy	p	Arg	Tyr	Cys	Gly	Ile	Val	Arg	Asn	Arg	Pro	Arg	Phe	Leu	Tyr	Gly	Arg	Ser	Lys	Ar	366		
70344	Db	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	70357		
366	Qy	g	Asn	Ala	Lys	Phe	Trp	Ser	Lys	Val	Ser	Ser	Lys	Lys	Lys	Lys	Met	Arg	Tyr	Gln	Al	386		
70358	Db	G	A	A	G	T	C	A	A	G	C	T	G	C	G	C	C	T	C	A	G	G	70417	
386	Qy	a	Thr	Ile	Gly	Leu	Gly	Ala	Leu	Ala	Ile	Ile	Leu	Leu	Tyr	Val	Ser	Leu	Arg	Phe	G	406		
70418	Db	T	G	C	A	T	A	G	G	A	T	G	T	G	C	C	T	T	G	G	C	C	70477	
406	Qy	u	Trp	Gln	Tyr	Ala	Phe	Ser	Ala	Val	Cys	Ala	Leu	Ile	His	Asp	Leu	Leu	Ala	Thr	Cys	Al	426	
70478	Db	G	T	G	A	A	T	T	C	C	G	T	G	G	G	C	G	A	T	C	C	G	70537	
426	Qy	a	Val	Leu	Phe	Ile	Ala	His	Phe	Phe	Leu	Lys	Ile	Gln	Ile	Asp	Leu	Gln	Ala	Ile	G	446		
70538	Db	T	A	T	C	C	T	G	-----	T	C	G	T	T	C	-----	C	A	G	A	T	C	70592	
446	Qy	y	Ala	Leu	Met	Thr	Val	Leu	Gly	Tyr	Ser	Leu	Asn	Asn	Thr	Leu	Ile	Ile	Phe	Asp	Arg	I	466	
70583	Db	G	G	C	G	T	G	C	C	A	T	C	A	C	C	A	C	A	C	C	A	C	70642	
466	Qy	e	Arg	Glu	Asp	-----	Arg	Gln	Ala	Asn	Leu	Phe	Thr	Pro	Met	His	Val	Leu	Val	482				
70643	Db	AC	G	T	GA	AA	AC	T	CC	G	CG	T	CA	T	GC	CA	AG	GC	T	CA	T	AT	70690	
482	Qy	l	Asn	Asp	Ala	Leu	Gln	Lys	Thr	Phe	Ser	Arg	Thr	Val	Met	Thr	Ala	Thr	Thr	Leu	Ser	502		
70691	Db	CA	AC	G	T	T	C	CC	CA	CC	AG	AC	CC	T	G	C	T	GC	GC	CA	CC	T	70750	
502	Qy	r	Val	Leu	Leu	Met	Leu	Phe	Ile	Gly	Lys	Ser	Ser	Val	Phe	Asn	Phe	Ala	Phe	Ile	Me	522		
70751	Db	GG	CA	T	GG	CC	GC	G	T	G	T	T	CT	TC	GG	GG	G	T	GA	AC	CT	G	70810	
522	Qy	t	Thr	Ile	Gly	Ile	Leu	Leu	Gly	Thr	Leu	Ser	Ser	Leu	Tyr	Ile	Ile	Ala	Pro	Pro	Leu	542		
70811	Db	GT	T	CA	T	T	GG	T	GT	CA	CT	AC	T	AC	T	CG	T	CG	A	T	AT	CG	70870	
542	Qy	u	Phe	Met	544	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	70871		
70871	Db	CT	GG	CT	G	70877	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	70877		

RESULT 14  
 LOCUS AE012971/c  
 DEFINITION Chlorobium tepidum TLS  
 ACCESSION AE012971 AE006470  
 VERSION AE012971.1 GI:21648307  
 KEYWORDS  
 SOURCE Chlorobium tepidum TLS  
 ORGANISM Chlorobium tepidum TLS  
 Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 Chlorobium.  
 REFERENCE 1 (bases 1 to 7687)  
 AUTHORS Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,  
 Dodson, R.J., Deboy, R.J., Gwinn, M.B., Nelson, W.C., Haft, D.H.,  
 Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,  
 Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,  
 Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,  
 Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,  
 Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.  
 TITLE The complete genome sequence of *Chlorobium tepidum* TLS, a  
 photosynthetic, anaerobic, green-sulfur bacterium  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)  
 REFERENCE 2 (bases 1 to 7687)  
 AUTHORS Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,  
 Dodson, R.J., Deboy, R.J., Gwinn, M.B., Nelson, W.C., Haft, D.H.,  
 Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,  
 Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,  
 Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,  
 Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,  
 Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-APR-2002) The Institute for Genomic Research, 9712  
 MEDICAL Center Dr, Rockville, MD 20850, USA  
 REMARK unpublished  
 FEATURES  
 source  
 1..7687  
 /organism="Chlorobium tepidum TLS"  
 /mol\_type="genomic DNA"  
 /strain="TLS"  
 /db\_xref="taxon:194439"  
 163..780  
 /gene="rnhB"  
 /note="synonym: CT2261"  
 163..780  
 /gene="rnhB"  
 /notes="identified by match to PFAM protein family HMM  
 PF01351"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribonuclease H1"  
 /protein\_id="AA073476.1"  
 /db\_xref="GI:21648308"  
 /translations="MLSTDFEYVWESLSQVCGIDAGRGPLAGVPVAAAVFPRHPR  
 PTGIFAKLDDSKLTAEIREDALAIRESAESALGVGDAETIDRINILQATMLANNL  
 AVESLGSTPEFLVDGNRRPRPVLPIPYQTIYKDGSKVFSIAAASVLAKTRHDELMTTS  
 AASYPIGFVEVHFGYPITARHVEAIAHRGCAIHRQSFKLKRGK"  
 800..1189  
 /gene="CT2262"  
 800..1189  
 /gene="CT2262"  
 /notes="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA073477.1"  
 /db\_xref="GI:21648309"  
 /translations="MAAPQWLGESEKIAARHLAAKGYRIVARNYRPHRNEIDIIAFD  
 GEALCEIVTKARSLGKRGHPASVTRSKQKEIAARAAAGYLASLDDPWITCRFDVIAVL  
 ALSIDERSINKYIEIHKAFAFMVGDGK"  
 1262..3205  
 /gene="gvrB"

```

CDS
/note="synonym: CT2263"
1262..3205
/feature="gyrB"
/note="identified by match to PFAM protein family HMM
PF00204"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="AAM73478.1"
/db_xref="GI:21648310"
/translation="MQETDIQTAQNASTEYGTATNIQVLDGIEHVRKRPAMYIGDIHR
GLHLVBIQVNAIDETLAGYNIHVAMNPDGVTVDHGRGIPVDIHPVKKSALE
LVMVTGAGGDFGAYKVGSLGCVASVNVALSCEVVEYRDGKIIFQTVRRGVP
QGVVEIGTDOGRKTVTKPDPEIKITEPRKIIDRMRELAFLANSRLRIIVODAE
QNEBIFHEGGLKEFVRFDNSRLSLKKEPIFLSGERDSTWETALQYDNOYNOIES
YVNNINHEGTHVTGRKALTRILNAYAKNDLLKNVKITLTDGDFKEGLTAVISVK
VPQFQGTQKLGNSGTSIVETI VNEQALDEFAENPGTLKLIIEKVSAAISREA
ARKAKETRRKSLGSLPGKADCSINDPDHCELYIVEGDSAGGAKOGRDRSFA
ILPKGILNVEKARLHMLNEIKITIIIALGTSIGEEFSPKLYRGKIIIMTDAF
VDGAHRTLLTFFFRMRSLEIAGKVFIAOPPLYLVKSGREQQYAWDEDERLSIVEQ
MKLKQKGANVSIORYKGLGEMNPEQLWSTMTDPAHRSLLQVTVENAMEADQVPSTLM
GDKVEPRDFIEKNARYVRLLDV"
/complement(3194..4510)
/gene="surA"
/note="synonym: CT2264"
complement(3194..4510)
/gene="surA"
/note="identified by match to PFAM protein family HMM
PF00639"
/codon_start=1
/transl_table=11
/product="peptidyl-prolyl cis-trans isomerase SurA"
/protein_id="AAM73479.1"
/db_xref="GI:21648311"
/translation="MKKVLFAVLAALMIAMNGFADAAASTGLDRIVAGNNEILASD
VNEQLMLHQPYETRPDQPKRLKLENNMINKIILTKAKIDTVKDEKVDQAAAR
YSSLRAGPVSAMESFRGPNRLKQHIRREDIRDOQMEAFRKNPHEVTVSYDET
AFYNOEGKALPEAPETVSQIIRKPLVSEAAAOALDKIKAVQOOLEAGSGPATLAR
EYSDPGSGRGGDLGRKGLVPSPEEAASVLKPGQISGIVETRGYHIIOLIDE
GDRITHILALPDRSKTIDPATIALKLSRKQVLSKATFAEMAKYISDDPASATNG
GLITSGNPDPAETLRPDLRKIDGLKSGDISQPEKIESKSPAFYALFMLNSRT
PAHVLTPHQAFOELALNHSQSELFNAWIEKLKREVVVKVMSDI"
/complement(4769..7544)
/gene="CT2265"
/note="this region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
GB:AE000513; identified by sequence similarity; putative;
protein-export membrane protein SecD, putative"
BASE COUNT 1761 a 2213 c 2162 g 1551 t
ORIGIN

Alignment Scores:
Pred. No.: 8.12e-57 Length: 7687
Score: 736.00 Matches: 184
Percent Similarity: 54.50% Conservative: 119
Best Local Similarity: 33.09% Mismatches: 202
Query Match: 26.49% Indels: 51
DB: 1 Gaps: 13

US-09-868-987-14 (1-552) x AE012971 (1-7687)

Qy 3 SerSerProLeuLeu-----AsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db AGCGCTCGGTCGTCGAGTCGCGCATTCGAGCGGCAAC---TCGGTCATCAATGGCATC 6327

Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db GGCAGCCTCGACGAGCCAAAGACCTCGAAATTATCTCAAGGCGGTGCTCTGCCCGCG 6267

Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuLysGlnCysThr 60
Db CCGGTGAAATTCCTCGAAGAGCGCACCGTCGCGCCGCTGCTTGGCGCGCACTACATCCG 6207

61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 80
6206 GCTGGTCTCTTTCTCTGATCTGGGTTCACGGCGTCTATCTTCATGATTGTTTAT 6147

81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr 100
6146 TATAACAAGCGGCGGTATCGCGCGACCTCGCGCTCGTCAACATTATGGTCATCATC 6087

101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
6086 GCCGTGCTTCGGGGTTACGGCGCTACCTACCTACCGTATTCGGGTATCGTCTG 6027

121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
6026 ACATGGGTATGCTGTTGACGCCAACGCTGCTGATTTACGAGCGCGTCCGCGGAGCTT 5967

141 LeuLeuSerGlnSerLeuLysSerValGluGlyGlyTyrThrLysAlaPheGlyAla 160
5966 GCCAGCGCAAGACCTGCGCATGCTGCTATCGACCGCGCTACAACCGTGGTTCATCG 5907

161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
5906 ATTCTGACTCGCAGCTGCGCGCTTGGCGCGCTTCTGCTCTATATTACGGCGTC 5847

181 GlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThr 200
5846 GGCCCAATTCAGGATTCGCCGTCACCTCATGCTACTCGCGGAGCTTGTTCAG 5787

201 AlaLeuPheMetThrLysPhePheMetLeu-----TrpMetAsnLys 215
5786 GCGATCGTCTGACCGAGGTGCTGTCGACCTGCTCGACAGCAAGTTTATGAATGAA 5727

216 -----ThrGlnHisThrGlnLeuHisMetMet-----AsnLysPheVal 228
5726 AAGAGTTTCGATTAATTAACAGACACTGCAATGCGCCTTTTCAGAAAAACAAATTCAA 5667

229 -GlyIleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerVa 248
5666 TTTTATCAAGCATCGC-----AGATCGCTATCGCATCTCTCTCGATTCT 5622

248 lPheLeuLeuGlyCysValAlaLeuLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAs 268
5621 GCTTGTGCGGCGCATGATCTCCCTGCTTCAAGGGAATGAAC-----TATGGCATCGA 5568

268 pPheLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAl 288
5567 CTTCAAGGCGGCTCGAGGTGCTCATCGCTTCGAGAAA-----GCAGCCGATGTC-- 5516

288 aGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPh 308
5515 -----GGCACCGTGGCTCGATCATCGAGACGCTGGAGTGAGCGGC----- 5474

308 eArgIleGlnThrPheGlySerSerGluLysIleTyrPheSerAspLysAlaLe 328
5473 -----TCGGTCAGGTCAGTAGTATGCGCATGCGATCGTTCGCT 5442

328 uSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTy 348
5441 GTTGCTCAAGACCGATTTCAGGGTGACACCCCGAGAGTCAAA----- 5399

348 rCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAl 368
5398 -----ACCTTGATTCTGAATCGCTGCGCTCGCGCAATCTCGCGCAGATTCT 5343

368 aLysPheTrpSerLysLysSerSerLysLysSerLysLysMetArgTyrGlnAlaThrIle 388
5342 GAGCGTC---GAGGCGGTCCGGCGGAGCATCGCTCCGACATGAAATATGTCGCCCATCAA 5286

388 eGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTrpG 408
5285 GGCACATCGCGCGCGCGCTGTCGCGCATCTGCTCTACGTGAGCGTGGCTTCGAGTTCAA 5226

408 nTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaVal-- 427

```



```
/note="similar to GP:5123490; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ompA family protein"
/protein_id="AA054691.1"
/db_xref="GI:28851615"
/translation="MFTSRRLIIIVATAVALLSGCATSNPNYDNOGAOSSGMSKTAKY
GGGLAGALAGAAIDHNNRKGALIGAAGVAGASAGVYADQEAALRASMANITGV
EVQGGQDIKLMFGNITFYATDSSAIASSFYSLPNLANSKQFNQNNIIEIGYDST
GSRQNMDSQQRAQSVATYLTSCQVDQSHLSVRGAGPDQPIASNADANGAQRNRVE
VNLKPIPGQQY"
6822..7922
/locus_tag="PSPT01165"
6822..7922
/locus_tag="PSPT01165"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA054692.1"
/db_xref="GI:28851616"
/translation="MFFFMPLPSWDSIRRLQDRDIAPYRGHFFTRPNEVPVSPALIM
RFGNVSFAFDRAWEACRIQLQRFVLDLSSIIKLDLIDVVTQAMTVVGSFAFVGGAA
GAGAVLFFGVGAVPGMLGAALGAQSAFIIIGLEIAEASVIDGPIRIEYTRG
ITTAWQGPREGQRPFGWDDLAQNSAVOEIAQGHVEVILLGAMVOYITRGRGNAR
ALANERASAKGERLPQGMLEHDELKRPDLORPEPRGALAEQETLSNRSKTADA
AKPDAMQKTVSSKIDPKQLDLVTGHNHSLSVHGGKVSVDGDLMLRAKSAIKPNQDV
GPSVPLSSAFHSDLELIFADQITRETLKVS"
7963..8940
/locus_tag="PSPT01166"
7963..8940
/locus_tag="PSPT01166"
/note="similar to GP:13094159, and GP:4514540; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ISPssy, transposase"
/protein_id="AA054693.1"
/db_xref="GI:28851617"
/translation="MKQMTFADAEYAGKRRHRRERFLIEMQVVPWKGLIALIEPHY
PKGGGPAYPLAMLRHLMQNFYSDPAMEALYETTLIRQFAGLSLDRIDETTT
ILNRLLEKHELAGGLQVINGVGLRGLMQGTVDVATIIHAPSSTKNKDKGRDP
EMHOTKNGNOYFFQMKHAGVDAESGLVSLVGTAAVADVTOVDOLHGEETVSGD
AGYTVGDKRAEHDROMWISIAAPSRVKHGEKSLIARVYRKIEFTKQALRAKVEHP
FRVTKRQFGYTKVPRGLAKNTAQQAQTLFALSNLNLMVWRKLLAMGEVRL"
10294..10887
/locus_tag="PSPT01167"
10294..10887
/locus_tag="PSPT01167"
/note="similar to GP:15024601; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA054694.1"
/db_xref="GI:28851618"
/translation="METFGQVQGRQEVPESSIDRYRGKLPNQLLTYWAEHGWCGYGDG
IFVTNQEYDVAVVAMLEGTKEFTFDAYHLIARSAFGDLYLWGEIEGASLKITSILS
RYTHKSLCIGSQMDKQLQSFILSRKVSNDYDGLFRPARKLGLTLHHDMEYGFVPAL
MLGGADTLDEKVKAVEHLVLLSQLSDLESYDFQDL"
complement(11408..12331)
/locus_tag="PSPT01168"
complement(11408..12331)
/locus_tag="PSPT01168"
/note="identified by match to TIGR protein family HMM
TIGR00530"
/codon_start=1
/transl_table=11
/product="acyltransferase family protein"
/protein_id="AA054695.1"
/db_xref="GI:28851619"
/translation="MDFLPAPLRGVIALSLLALNTIVCCTVLPFAVTLKICLPFSAAQ
```

```

RFTDMLMHIHETWIGTNNAWIRLCHTRWQLSGLGLDYKHSYLVTSNHSQSWVDIMV
LOVLANRIRPLKFLKELIIVPVLIGLAWALGFPMKRYSKAYLAKHPEKKGKOLE
TTTRTCAKPRHNVPVGIENAEGRFTAGKHQQOSPPRHLKPKAGGIAFVLDMGSEQ
LESIIINTIHYGGQFGYDMLCGNVTDVVAHFEELRIPQFLUGKNYDQGEYRLFEQ
QWINALWEKDSLLDLRLEHQYPAQRGRDAF"
complement(12396..14486)
/gene="pta"
/locus_tag="PSPT01169"
complement(12396..14486)
/gene="pta"

Alignment Scores:
Pred. No.: 1 08e-54 Length: 311249
Score: 736.00 Matches: 183
Percent Similarity: 54.06% Conservative: 123
Best Local Similarity: 32.33% Mismatches: 182
Query Match: 26.49% Indels: 78
DB: 1 Gaps: 16

US-09-868-987-14 (1-552) x AE016860 (1-311249)
```

```
Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db 284316 ATCATCAGCCTGGCGACCATCCAGTCGCCGCTGGCGAATCAGTTCCGGATTACCGGTCTG 284375
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 284376 AATGGTCAGGGTGAGTCTCCGAGCTGGCGCTCTGTTGCGCGGTGGTGTGGCGCGC 284435
Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
Db 284436 CCGATGTACTTCGCTGAAGAGCGCACCATCGGTCCGAGCCTGGTCCCGACACATCACC 284495
Qy 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 80
Db 284496 AAGGGTATCGACGCTTCCTCGTGGGCGATGCTGCTGATTCATCTTCATCGGCGATC 284555
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr 100
Db 284556 TACCGGTTCTTCGGTCTGATCGCTACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284615
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 284616 GCGCTGATGTCTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 284675
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140
Db 284676 ACCATGGGTATGGCGGTGGATGCCACGTGTTGATCTTCTCGCGGATCCGCTGAAGATC 284735
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 284736 GCCAATGGGATGACGGGTGCAACGGCTATCAACAGAGGTTTCGATCGCTACACGGCG 284795
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 284796 ATCTCGACGCCCACTGACGAGCTGCTGGTGGCGGGATCTCTTTCCGATGGGCACC 284855
Qy 181 GlyProIleGlyGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPheThr 200
Db 284856 GGCCCGCTCAAGGGTTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 284915
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220
Db 284916 GCATCATGTGACCGCGC-----GCTCGTGACTTC-----AGAAG 284933
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db 284934 ---GCTATGTCACCTGATCTACGGC---GGTCTGACTTC-----AGAAG 284975
Qy 241 LeuTyr-----AlaValSerGlySerValPheLeuLeuGly-----252
Db 284976 TTGTGGATTAAAGGGGCTGCCATGTGTACGTACCATCACTTCTTCATGGCGCTTCCCAATGTT 285035
```

```
QY 253 -----CysValAlaLeuGlyPheGlyAlaTrp-----Asn 262
Db 285036 GCGTTCCGCAATCACCAGTCCTTACCGCAGCTGGCTGTTCCAGCTGTTCCATAAGGG 285095
QY 263 SerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLysGluHis 282
Db 285096 CTCAACTTTGGTCTGGACTTACCGCGGCTACGCTCATCGAGCTGACCTACGAGCGCTCT 285155
QY 283 GlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGly 302
Db 285156 -----GCCGATCTG-----GCCAAGGTTCTGCAGGAGCTGTTTCGGCCCGGC 285197
QY 303 LeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleTyr 322
Db 285198 TAC-----CACGAAGCTGTTGTGCAGAGCTTTGGCGCACGACTGACTTGTGTGTGCGC 285251
QY 323 PheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr 342
Db 285252 ATGCCGGGTGAAGACCCGCAATTGGGCACCTCAGGTCTGCTGAAGCGCTGCGCAAG----- 285305
QY 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362
Db 285306 -----GCCGTGCCGATACCCCGCAGTGTGTC----- 285332
QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382
Db 285333 -----AAGCGTGTGAGTTTC-----GTCCGCCCGCAGGTCGGCGAAGAGCTG 285374
QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuTyrValSer 402
Db 285375 CGTGATCAGGGTGGCTCGGCATGCTGATGGCGTGGCGGTGTCTGATCTACTCCTGCGCG 285434
QY 403 LeuArgPheGluTyrGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422
Db 285435 TTCCGCTTCCAGTGAAGTTTCTCTCGCGCGATCATTTCTGCTGATCCACGACGTGTGTC 285494
QY 423 AlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeu 442
Db 285495 GTGACGATGGGTATCCTG-----TCGTTCTTC-----CAGATTACCTTCGACCTG 285539
QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIle 462
Db 285540 ACGGTATTGGCGCGGTGCTGCGCATCATCGGTACTCTGCTGAACACACCATTTGTGTC 285599
QY 463 PheAspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMet 478
Db 285600 TTCGACCGGGTTCGGAACACTTCGCGCTGCTGCGCAAGGCTTCGCTGATCGAGAAC--- 285656
QY 479 HisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAla 498
Db 285657 -----ATCAACATCTCCACACCGCAGACACTGCTGCGCACCATTTGCCACCTCGGTT 285707
QY 499 ThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPhe 518
Db 285708 TCGACCTCTGTTGGCGATCTCTGCACTGTGGCGGTTCTGGTGGCGCAGCCTGGAAGGCTTC 285767
QY 519 AlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaPro 538
Db 285768 TCCATTGGTGTTCATCGGTGTGTCGGGGTACCTACTCTGATCTATCATCGCCAAC 285827
QY 539 ProLeuLeuLeuPheMet 544
Db 285828 GTGGTGTGATCTGGCTG 285845
```

Search completed: October 28, 2003, 17:40:24  
Job time : 7083 secs

**THIS PAGE BLANK (USE70)**

```

FT XX                               /*tag= a
DN WO200039158-A1.
XX PD                               06-JUL-2000.
XX PF                               99WO-CA01230.
XX PR                               98US-0113280.
XX PR                               98US-0113281.
XX PR                               98US-0113282.
XX PR                               98US-0113283.
XX PR                               98US-0113284.
XX PR                               98US-0113285.
XX PR                               98US-0113285.
XX PR                               98US-0113385.
XX PR                               98US-0114050.
XX PR                               98US-0114050.
XX PR                               98US-0114056.
XX PR                               98US-0114057.
XX PR                               98US-0114058.
XX PR                               98US-0114059.
XX PR                               98US-0114059.
XX PR                               98US-0114061.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX XX
XX PI Murdin AD, Oomen RP, Wang J;
XX XX
XX DR WPI; 2000-452369/39.
XX DR P-PSDB; AAY95543.
XX PT Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
XX PT prevention and treatment of Chlamydia infection in mammals -
XX PS Claim 2(a); Page 55-58; 215pp; English.
XX CC
XX CC The present sequence is that of Chlamydia pneumoniae genomic DNA
XX CC including an open reading frame that codes for CPN100686 RY 54 (see
XX CC AAY95543), a putative 93 kDa outer membrane protein. It is an example
XX CC of C. pneumoniae polynucleotide molecules of the invention (see
XX CC AAY50030-42) that encode antigenic polypeptides (see AAY95543-55) useful
XX CC in the diagnosis, treatment and prevention of Chlamydia infection.
XX CC The polynucleotides can be utilised: in the recombinant production
XX CC of Chlamydia antigens using transformed unicellular host cells; in
XX CC vaccines or live vaccine vectors; in naked form or formulated with
XX CC a delivery vehicle for therapy and prophylaxis of Chlamydia
XX CC infection; and as probes or primers for diagnosis of Chlamydia
XX CC infection.
XX SQ Sequence 1864 BP; 498 A; 346 C; 401 G; 619 T; 0 other;

Alignment Scores:
Pred. No.:      2 27e-301      Length:      1864
Score:          2778.00      Matches:      552
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            21      Gaps:      0

US-09-868-987-14 (1-552) x AAAS0030 (1-1864)
Qy 1 MetValSerSerProileLeuAsnValProleuLysAsnHisAlaSerValSerGlyLys 20
Db 101 ATGTCAGCAGCCCTATTTTAAACGTCCTCCATTGAAATAATCATGCCAGTGCTCAGGAAA 160
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 161 TTTACCCACCGTGAAGTGAAGAGCAACTCGCCTCAGATTAAATCATGAGGCGATGCTCTTT 220
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
Db 221 GTTCCCGAGGTTCTCAGTGAAGAGACGACATCTCTCTGATCTTGGGAAACAAATGATACA 280
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
Db 1361 CTTTGGCTACCTGTGCGAGTCTTGTATAGCACATTTCTTTTGAAGAAATTCAAATA 1420

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:11 ; Search time 418 Seconds  
(without alignments)  
3564.808 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKHNASVSGK.....SLYIAPILLFMVRKNSK 552

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/Cyn2\_1/USPTO\_spool/US09868987/runat\_28102003\_152601\_23827/app.query.fasta\_1.711  
-DB=N Geneseq\_19Jun03 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987 -CGCN\_1\_0 @runat\_28102003\_152601\_23827 -NCPUP=6 -ICPU=3  
-NO\_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

N Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2778	100.0	1864	21	AAA50030
2	2730	98.3	1230025	20	AAX91990
3	1920	69.1	1038602	20	AAX91990
4	719	25.9	1830121	17	AAT42063
5	686.5	24.7	45613	22	AAF28535
6	678.5	24.4	2214	20	AAX29477
7	672	24.2	2943	20	AAX91694
8	672	24.2	2955	20	AAX91570
9	664	23.9	1584	24	ABK74549
10	653	23.5	2292	25	ABT15148
11	646.5	23.3	2226	22	AAX52892
12	646.5	23.3	2301	24	ABN91389
13	646.5	23.3	3511	22	AAX54037
14	646.5	23.3	3705	22	AAX54228
15	618.5	22.3	349646	21	AAA81457
16	618.5	22.3	349980	21	AAF21607
17	618.5	22.3	349980	21	AAF21608
18	618.5	22.3	1437668	21	AAA81490
19	618	22.2	495269	24	ABQ67195
20	618	22.2	3011208	24	ABQ69245
21	616	22.2	2944528	24	ABA03041
22	612	22.0	5839	24	ABQ70971
23	589	21.2	910715	20	ABX20248
24	512	18.4	4403765	22	AAI99683
25	509	18.3	80557	24	ABX09142
26	509	18.3	4411529	22	AAI99682
27	461.5	16.6	18477	23	AAS59634
28	429	15.4	349980	22	AAH68529
29	414	14.9	835	22	AAF26302
30	408.5	14.7	1848	22	AAF94404
31	408.5	14.7	1848	24	ABK64968
32	393	14.1	1671	19	AAX13967
33	386.5	13.9	1581	18	ABQ25997
34	381	13.7	1848	24	ABQ90076
35	368	13.2	1024	24	ABX65642
36	368	13.2	1024	24	ABX66280
37	368	13.2	1024	24	ABX66285
38	365	13.1	1377	20	AAX20685
39	358	12.9	2322	23	AAS93101
40	357.5	12.9	2323	23	AAS89009
41	357.5	12.9	2326	23	AAS90384
42	350	12.6	1035	19	AAX14487
43	346.5	12.5	1854	25	ABZ38871
44	346.5	12.5	1857	24	ABK54085
45	342	12.3	975	22	AAF94409

## ALIGNMENTS

RESULT 1  
AAA50030  
ID AAA50030 standard; DNA; 1864 BP.  
XX  
AC AAA50030;  
XX  
XX  
XX 10-OCT-2000 (first entry)  
DE  
DE DNA encoding Chlamydia pneumoniae antigen CPN100686 RY 54.  
XX  
XX CPN100686 RY 54; antigen; infection; diagnosis; therapy; vaccine;  
XX outer membrane protein; ds.  
XX Chlamydia pneumoniae.  
XX  
XX Key Location/Qualifiers  
FT CDS 101..1759



Qy 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeu 460  
 Db 1421 GATTGGCAAGCCATTGGTGCTTTAATGACTGATTTGGGTATTTCATTAACAACACTTTG 1480  
 Qy 461 IlellePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHisVal 480  
 Db 1481 ATCATTTTGTATGCTGATTCGTGAAGATCGCAAGCGAACCTGTTTACCCCTATGCTGTT 1540  
 Qy 481 LeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThr 500  
 Db 1541 TTAGTTAATGATGCCCTTCAAAGACGTTACGCCGACGGTAATGACAAACAGCTACAACT 1600  
 Qy 501 LeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAlaPhe 520  
 Db 1601 CTATCAGTTTGTGTAATGCTTTGTTATAGCGGCTCTCTGCTTTAATTTGCAATTT 1660  
 Qy 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeu 540  
 Db 1661 ATTATGACCATAGGATTCCTTAGGAACCTTATCGTCTCTTTATATGACCACTCTG 1720  
 Qy 541 LeuLeuPheMetValArgGlyGluAsnArgSerLys 552  
 Db 1721 TTGTTGTTTATGTCCTGTAAGAAATCGCTCAAAA 1756

## RESULT 2

AAX91990/c  
 ID AAX91990 standard; DNA; 1230025 BP.

AC AAX91990;

XX 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.

XX Chlamydia pneumoniae.

XX W09927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Claim 1; Page 291-611; 1912pp; English.

XX The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

## Alignment Scores:

Pred. No.:	1.38e-291	Length:	1230025
Score:	2730.00	Matches:	551
Percent Similarity:	99.28%	Conservative:	0
Best Local Similarity:	99.28%	Mismatches:	1
Query Match:	98.27%	Indels:	3
DB:	20	Gaps:	0

US-09-868-987-14 (1-552) x AAX91990 (1-1230025)

Qy	1	MetValSerSerProfileLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys	20
Db	661906	ATGTCAGCAGCCCTATTTTAAACGTCCTCCATGAAATCATGCCAGTGTCTCAGGAAA	661847
Qy	21	PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe	40
Db	661846	TTTACCACCCGTTGAAGTGAGCAAACTCGCTCAGATTTAAATCTGGAGCGATGCTTTT	661787
Qy	41	ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr	60
Db	661786	GTTCCCGAGGTTCTCAGTGAAGAGCGATCTCTTCATCTTGGGAAAAACAATGTACA	661727
Qy	61	GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr	80
Db	661726	CAAGGCATTATCTCAGCATGCTGTGGCTTGCAATGCTTATTTGTCATGAGCGTATAT	661667
Qy	81	TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp	100
Db	661666	TATAGATTGGAGCGCTCATCGCTTCGGAGCTGCTCTTCGAAATCTTTTCTTATCTGG	661607
Qy	101	AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu	120
Db	661606	GCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTCCAGGACTCGCTGGGATGTTCTT	661547
Qy	121	AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe	140
Db	661546	GCTATGGGATGGCGGTAGATGCAAAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTT	661487
Qy	141	LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla	160
Db	661486	TTATTTGCTCAAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC	661427
Qy	161	IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr	180
Db	661426	ATTTTTCATCTTAACCTTGACTACAGTATTTGGCCTCAGCACTCTCTTTCTTCTTAGATACA	661367
Qy	181	GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr	200
Db	661366	GGGCTATTAAAGGGTTTGCTTTGACATTTAGGAATTTTCTCTTCAATGTTTACG	661307
Qy	201	AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln	220
Db	661306	GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG	661247
Qy	221	LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys	240
Db	661246	TTGCATATGATGAATAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGGATGCACAAAAA	661187
Qy	241	LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla	260
Db	661186	CTTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTAGTTGCGTGTGCTCTCGGTTTGGAGCC	661127
Qy	261	TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys	280
Db	661126	TGGAAATTTCCGTTTGGGAATGGATTTTAAAGGAGGGTATGCCCTTTTAAATCCAAAA	661067
Qy	281	GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu	300
Db	661066	GAGCATGCGATCAGCGATGTTGCTCAATCGTGGCAAAAGTTGTGTCATTAACACTACAGAA	661007
Qy	301	AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys	320
Db	661006	GCTGGTCTTTCTCTAGAGACTTCCGTAATTCGATTTCAACATTTTGGATCTTCAGAAAAAGATCAAA	660947



```
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyVilePheSerSerMetPheThr 200
Db 992651 GGACCTATCAAGGTTTCGCGCTCACTTTGATCATCGGAGTTTCTCATCATGTTTCA 992710
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220
Db 992711 GCCCTGTTTATGACAAATTTTCTCTTTGCTGTTGCGTTCGAGAAACGAGAGACCCAA 992770
Qy 221 LeuHisMetMetAsnLysPheValGlyLysHisAspPheLeuArgGlyCysLysLys 240
Db 992771 TTGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 992830
Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 992831 TTGTGGGTAGTTTCTGGAGTAGTTCTCTTTAGGTGTTGAGGATTAGGTTTGGTCT 992890
Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyLysThrAlaPheThr----- 276
Db 992891 TGGGATCCGTTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 992950
Qy 277 -----PheAsnProLysGluHisGlyLysSerAspValAlaGlnMetArgGlyLys 293
Db 992951 GTGTGTGAGTCAATCCAGAG-----CAGATGTGTCTGTA 992986
Qy 294 ValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgGlnThrPhe 313
Db 992987 TTGAGAAAGCGTTTCCACAAATCGGATATCTCTCGAGATTATCGTGTGCGTAGAGCA 993046
Qy 314 GlySerSerGluLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGln 333
Db 993047 GATAGTTTCTGGAAAGTCAAAATATATTTATCTCAGAAATGCTTGGATAGAGTTGAGCAA 993106
Qy 334 IleArg-----AlaSerLeuLeuLys 340
Db 993107 ATAGAAAGAGCTGGATCAGACAGAAAGATCGGATATCATCTTGTCTGAGTCTTCAA 993166
Qy 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValAlaGlnAsnArgProArgPheLeu 360
Db 993167 GTATTGTCTGATTCC----- 993181
Qy 361 TyrGlyAsnSerLysArgAsnAlaLysPhe-----TrpSerLysVal 374
Db 993182 ---GGAAGTTCTACTACTTCTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT 993238
Qy 375 SerSerLysLeuSerLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeu 394
Db 993239 AGTGACAGCTTTTCCATAAATCGTACACAGGCTGTCATCGCATTTATTTGGAGCTTTA 993298
Qy 395 AlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaVal 414
Db 993299 GGAATTATCTTATTTATGTTAGTTTGGCTTTGAGTGAGATATGCTTCAGTCCGAT 993358
Qy 415 CysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePhe 434
Db 993359 TGTGTTTAAATGACGAGCTTATAGCGACTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 993418
Qy 435 LeuLysLysIle-GlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyr 454
Db 993419 TTGCAGAGATTACAGATTGATTTCACAGCAATAGGCGCATTTGATGACAGTCTAGGGTA 993478
Qy 454 rSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLe 474
Db 993479 CTCTTTGACAAATACGTTGATTTATTTTGTGCTGATTCAGAGATCGCGCGGAAAAACT 993538
Qy 474 uPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVa 494
Db 993539 GTTCACCCCAATGCGGATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 993598
Qy 494 lMetThrThrAlaThrLeuSerValLeuLeuMetLeuLeuPheIleGlyLysSerSe 514
Db 993599 TATGACTACGGCAACAACTCTATCTGCTACTGCTATCTTGTGTTTATTTGTTGTTGTTGTT 993658
```

```
Qy 514 rValPheAsnPheAlaPheIleMetThrIleGlyVileLeuLeuGlyThrLeuSerSerLe 534
Db 993659 TATTTTCAATTTTGCATTTCATTATGACAGTGGGATCTTGTAGGAACGCTATCTTCATT 993718
Qy 534 uTyrIleAlaProProLeuLeuLeuPheMetValArgLysGluAsnArgSer 551
Db 993719 GTACATAGCTCTCCGCTCTCTTATTTCATGCTGCGTAAAGGAAGACAAAT 993770
RESULT 4
AAT42063/C
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
XX 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
XX Genome; bacterium; Haemophilus influenzae; computer readable medium;
XX KW expression modulating fragment; regulation; gene expression; vector;
XX KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN W09633276-AL.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
XX 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX WPI; 1996-485782/48.
XX
XX Haemophilus influenzae Rd genome recorded on computer readable
XX medium - useful for identifying commercially important nucleic acid
XX fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
XX This sequence represents the complete genome sequence of the bacterium
XX Haemophilus influenzae strain Rd. The invention relates to a computer
XX readable medium (CRM) having recorded upon it the complete H. influenzae
XX nucleotide sequence (I), a representative fragment of (I) or a nucleotide
XX sequence at least 9% identical to (I). By providing the full-length
XX genomic sequence in a computer readable form, it is possible to identify
XX commercially important nucleic acid fragments and expression modulating
XX fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
XX regulate the expression of a nucleic acid molecule. Vectors and altered
XX organisms comprising the predicted ORFs can be used to produce any of the
XX polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
Alignment Scores:
Pred. No.: 1,78e-65 Length: 1830121
Score: 719.00 Matches: 186
Percent Similarity: 55.69% Conservative: 127
Best Local Similarity: 33.10% Mismatches: 174
Query Match: 25.88% Indels: 75
DB: 17 Gaps: 18
US-09-868-987-14 (1-552) x AAT42063 (1-1830121)
Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22
||| |||:||||| |||:|||||
```



PT acids -  
 XX PS Claim 1, Page 180-191, 545pp; English.  
 XX  
 CC The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.  
 XX  
 SQ Sequence 45613 BP; 13249 A; 9346 C; 9947 G; 13071 T; 0 other;

## Alignment Scores:

Pred. No.: 2.53e-64 Length: 45613  
 Score: 686.50 Matches: 178  
 Percent Similarity: 50.5% Conservative: 126  
 Best Local Similarity: 29.5% Mismatches: 191  
 Query Match: 24.71% Indels: 107  
 DB: 22 Gaps: 14

US-09-868-987-14 (1-552) x AAF28535 (1-45613)

QY 17 ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGly 36  
 DB AATCAGGCTTAGATTAGATGCGAAGCAGCGGCTTGGCTTACTGTAGCTCAGGC 28831  
 QY 37 AlaMetSerPheValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLys 56  
 DB GCGTTAGCGGCACCGCATGCTTCTTGAAGAGCGCAACCATTTGGCGCATCAA 28771  
 QY 57 LysGlnCysThrGlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuValLeu 76  
 DB GATAATATCGATAAAGGATTATCTCTACACAGGTGGCTATCTGTGTATTTATGG 28711  
 QY 77 MetSerValTyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeu 96  
 DB ATGATTGTTTTATCGTCATTTGGTGTGATGCAAAATATGCTTTAGCCCTTAATGTC 28651  
 QY 97 LeuLeuIleTrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAla 116  
 DB ATTATCATGCTGGTGCATTATGTCCTTATGATCATCGCTCACCTACCTGGTATTGCT 28591  
 QY 117 GlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIle 136  
 DB GGGATTGTTGACCATTTGGTATGCGGTTGATGCAATGCTGAATTTTGGCGGTATT 28531  
 QY 137 ArgGluGluPheLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLys 156  
 DB CGTGAAGAGCTTGCCAAATGGTACTCGTGCAAAATCCGCCATTTGGCGAGGTTTCATCGT 28471  
 QY 157 AlaPheGlyAlaIlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuPhe 176  
 DB GCATTAGCTCAATTTTGTATGCCAATATCACAGCTATTAATTCATTATCTATT 28411  
 QY 177 PheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSer 196  
 DB GCCATCGGAACAGGACCAATCAAGGTTTGGCAATCAGCTTGCATATGATTATCAGT 28351  
 QY 197 SerMetPheThrAlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThr 216  
 DB TCTTTGTTTACTGCGATTATTAGTCACTCGTGCCATGATTCAAAATTTGGTATGGTCATCGC 28291  
 QY 217 GlnHisThr-----GlnLeuHisMetMetAsn----- 225  
 DB AAAAATTAACCAAACTTAGCATCGTTAGGAGTTAAAAATGAATAAATAAACCTT 28231  
 QY 225 ----- 225

DB 28230 ATCGACCAAATGCGACGACAGCCACCACGCGCTGTCGCGCTGTGATGGTAAA 28171  
 QY 225 ----- 225  
 DB 28170 GGTGCCAATGCTGCCACAAATCAAGATGCTCTTCCGATCAATCATTTGTCGACGCT 28111  
 QY 226 -----LysPheValGlyIleLys-----His 232  
 DB 28110 GGTGATGAAGCTGCGCAAAAGCGGTGGTGTCAAACTCATCGCTAATCATCGGTATCAT 28051  
 QY 233 AspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGly 252  
 DB CCTTTTATGAAGCTTGAAGCCCATGATAGTATTTTCGGTATTGCTCTATATCATAGG 27991  
 QY 253 CysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGly 272  
 DB CTGATAGCAATCATTTGTCAGGGTTTAAAC-----TTTGGTTGGATTTTACAGGTGCT 27937  
 QY 273 TyrAlaPheThrPhe-----AsnProLysGluHisGlyIleSerAspValAlaGln 289  
 DB GTGTGAGCAATGCTGCTTATGAACAGCGCTGCTTCAG----- 27901  
 QY 290 MetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArg 309  
 DB ---CAAGCACAAATCATTAAGCATTTGGCTGACAATGCTTTAAT-----GATCGCGTT 27850  
 QY 310 IleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAlaLeuSer 329  
 DB GTCCAAATACCTAGGAACCTGCCAAGAGCTTTGGTAGCTTACCGCCCAAGATGGCTCA 27790  
 QY 330 TyrThrLysGlnIleArgAlaSerLeuLysLeuThrIleMetSerTrpArgTyrCys 349  
 DB ---GCTGAAATCTAAGTGCTTCTTTGGATCAG----- 27760  
 QY 350 GlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLys 369  
 DB -----ATTCTTAATTTTCCA-----AATAATCCATCAACCATTCAT 27724  
 QY 370 PheTrpSerLysValSerSerLysLysLysMetArgTyrGlnAlaThrIleGly 389  
 DB AGCATTAGCATTTGCGCAGCCAACTGGCAATGAATTTACCTAAATTCATCGCTCGG 27664  
 QY 390 LeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyr 409  
 DB ATGCTGTGTCATTTGGTATGATGCTCATTTATGTGCAACACGATTTCCAGTTCAGCTG 27604  
 QY 410 AlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPhe 429  
 DB GCTTTGGTGCGGTACTTGTCTTTATTTTCATGATG-----ATTATCGTT 27559  
 QY 430 IleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMet 449  
 DB GTTGGTATTTTGGCATCATCGGCTGGCCATTTGACTTTGACGGTTTTGGCAGCTGTATGT 27499  
 QY 450 ThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp 469  
 DB GCTTTGATGCTATTCGTCGAATGATACATCGTTGTTTATGATCGTATTCGTCGAGAAT 27439  
 QY 470 ---ArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLys 488  
 DB TTTGCCGAGTGGCTGGTTTAAACACCCCGTCAGGTGCTT---GATTTATCACTACACAGA 27382  
 QY 489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 DB ACCTTAAGACCAACCATTTATGCTGTCGTCAGGTGTTTGTGTGTTGTTGTCCTATCGCT 27322  
 QY 509 PheIleGlyLysSerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 DB TTCTTAGGTGCTGATGCTGCTGTTGTTCTCTGTCGCGTTTATTATCGCTGATTCCT 27262  
 QY 529 GlyThrLeuSerSerLeuTyrIleAlaPro-----ProLeuLeuLeuPheMetValArg 546  
 DB GGTACATATCTTCCATTATGTTGCAAGCTCTATTCGCTGATGATGGTCTATCAAGA 27202

Alignment Scores:		
Pred. No.:	1.8e-65	Length:
Score:	678.50	Matches:
Percent Similarity:	50.25%	Conservative:
Best Local Similarity:	29.01%	Mismatches:
Query Match:	24.42%	Indels:
DB:	20	Gaps:
US-09-868-987-14	(1-552) x AAX29477	(1-2214)

QY 323 PheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr 342  
 Db 1633 TTTAAAGACAAA-----  
 QY 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362  
 Db 1645 -----TACGGA 1650  
 QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382  
 Db 1651 TCTGATCCAAATGTCAGACAGTTTCACCGACAGTCGGTAAGGAGCTGGCGAAGATCGG 1710  
 QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSer 402  
 Db 1711 CTGTACGCGAGTTGCTATAGCTTCTATTGGC-----ATCATTATTACGTTTCA 1758  
 QY 403 LeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422  
 Db 1759 ATCCGATTGGAATACAAAATGCGGATTGCTGCCATCGCTCATCTATATGAC----- 1812  
 QY 423 AlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeu 442  
 Db 1813 -----GCATTTCTTATGCTCAGCTTCTTACGATTATACAGGCTTGAGGTAGATGTT 1863  
 QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIle 462  
 Db 1864 ACATTATCTCGCGCCATCTTGACGATAATCGGCTATTCATTAAAGTACATACTGTACA 1923  
 QY 463 PheAspArgIleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHis 479  
 Db 1924 TTTCAGAGGTCGCGGAGCATATGAAAAGCGTAAGCGCGAAACCTTTGCCGATCTGAAC 1983  
 QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThr 499  
 Db 1984 CATATTGTAAACTGAGCTGAGCTGAGCAACCTTTACAGTTCAATTAACTGATTAAACC 2043  
 QY 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnPheAla 519  
 Db 2044 GTTGTGATTGTTGTGTGACATTGCTGATCTTTGGAGCATCTTCTATTACTAACTTCTCA 2103  
 QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
 Db 2104 ATTGCTTTATTGCTGGGCTGTTAACAGCGCTTTATTCTTCTATACATTGCCGCACAA 2163  
 QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552  
 Db 2164 ATTTGGCTTGC-ATGGAAGGAGAGAACTGAAAAAGA 2201  
 RESULT 7  
 ID AAX91694  
 XX AAX91694 standard; DNA; 2943 BP.  
 AC AAX91694;  
 XX  
 XX 20-MAR-2003 (updated)  
 DT 25-AUG-1999 (first entry)  
 XX  
 DE Porphyromonas gingivalis protein PG18 encoding DNA.  
 XX  
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 KW vaccine; antigenic; ds.  
 XX  
 OS Porphyromonas gingivalis.  
 XX  
 XX WO929870-Al.  
 PN  
 PD 17-JUN-1999.  
 XX  
 PF 10-DEC-1998; 98WO-AU01023.  
 XX  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003128.  
 PR 05-MAY-1998; 98AU-0003338.  
 PR 22-MAY-1998; 98AU-0003654.  
 PR 29-JUL-1998; 98AU-0004917.  
 PR 30-JUL-1998; 98AU-0004963.  
 PR 04-AUG-1998; 98AU-0005028.  
 XX (CSLC-) CSL LTD.  
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rachel LJ, Webb EA;  
 XX WPI: 1999-385613/32.  
 DR P-PSDB; AAY34476.  
 XX Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 XX Claim 12; Page 202-203; 588pp; English.  
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 CC (Updated on 20-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 2943 BP; 704 A; 696 C; 719 G; 824 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,5e-64 Length: 2943  
 Score: 672.00 Matches: 189  
 Percent Similarity: 51.03% Conservative: 109  
 Best Local Similarity: 32.36% Mismatches: 213  
 Query Match: 24.19% Indels: 74  
 DB: Gaps: 18  
 US-09-868-987-14 (1-552) x AAX91694 (1-2943)  
 QY 3 SerSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGlyLysPhe 21  
 Db 1300 TCTGCTCCGAACGTGAATGATGAGATCAGCGGCGCTCTCTCAGATCTCCGGGCACCTTC 1359  
 QY 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 1360 ACCGTGGAGGCGCGGTGACCTTGCCACGATCTCACTCCGTAATAATGGATGCTACG 1419  
 QY 42 ProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61  
 Db 1420 GTAACGATCGAACAGGAAACGTCATTGCTCTACGCTGGGTGCGAGTCCATTAAAGCC 1479  
 QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 81  
 Db 1480 GGATTCTTGTGCTCTCTGCTGCTTGTGTTATCTCTGATGTTTACATGTGCTCGCTTAC 1539  
 QY 82 ArgPhe---GlyGlyValIleAlaSerGlyValValLeuLeuAsnLeuLeuIleTrp 100  
 Db 1540 GGTTCCTTCGCGGCTCTTATCCGAAACGGCGCATGATTGTAACACGCTTCTTCACATTG 1599  
 QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 1600 GCGTATTGGCTCTTCTCCATGCTGCTGACCTCTCGGCTATCGCAGGTTTGGTGTGCTG 1659  
 QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
 Db 1660 ACGTGGGTATGGCTGGATGCCAACGATCTTATCTTCGAGCGGTATCAAAAGAGAGCTT 1719







CC assays. Porphyromonas gingivalis is involved in periodontal disease especially gingivitis.  
 CC (Updated on 20-MAR-2003 to correct PR field.)  
 XX

SQ Sequence 2955 BP; 711 A; 698 C; 721 G; 825 T; 0 other;

## Alignment Scores:

Pred. No.: 1, 516-64 Length: 2955  
 Score: 672.00 Matches: 189  
 Percent Similarity: 51.03% Conservative: 109  
 Best Local Similarity: 32.36% Mismatches: 213  
 Query Match: 24.13% Indels: 74  
 DB: 20 Gaps: 18

US-09-868-987-14 (1-552) x AAX91570 (1-2955)

```

QY 3 SerSerProLeuLeuAsnValProLeuLys---AsnHisAlaSerValSerGlyLysPhe 21
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1312 TCTGCTCCGAACGTGAATGATGAGATCAGCGCGGTCTCTCAGATCTCCGGGCACTTC 1371
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1372 ACCGTGGAGAGCGCGGTGACCTTCCCAACGTACTCAACTCCGGTAAATGATGCTACG 1431
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThrGln 61
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1432 GTAACGATCGAAGCAAGAAACGTGATTGGTCTACGCTGGTCCGAGTCCATTAAGCC 1491
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTrp 81
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1492 GGATTCCTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 82 ArgPhe---GlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1552 GGTTCCTTCCCGGCTCTTATCCCAACCGCGCATTTGATTGTAACAGCTTCTTCACATTG 1611
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1612 GCGGATTGGCTCTTCCATGCGGTGCTGACCTCTCGGTATCGAGGTTTGGTGTG 1671
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1672 ACGCTGGGTATGGCTGTGGTCCCAACGTACTTATCTTCGAGCGGTATCAAGAAGAGCTT 1731
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1732 CGTCCGGTAAACCTCCGATTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 1791
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1792 ATCTTCGATCGAAGCTTACGACTATTATTACCGGTATCATCTTCTCTACGGGAGC 1851
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 181 GlyProIleLysGlyPheAlaLeuThrIleLeuLeuGlyIlePheSerSerMetPheThr 200
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1852 GGGCGGATTCCGGGTTTGGCCACTACGTGATTATCGGTCTTATCGGTCTTCTTATTACG 1911
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 201 AlaLeuMetThrLysPhePhePhePheMetLeuTrpMetAsnLysThrGlnHisThrG1 220
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1912 GCTGCTCTTCTGACTCGTATCGTCTCGAGAACTGGCGAAAGGTCGTTGGATAGG 1971
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 220 nLeuHisMetMetAsnLysPheValGlyIleLys-----HisAspPheLeuAr 236
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1972 ATTACATTCACGAGCATTTACTCGCAATCTCTTGTCTCAATCCCTCATACAC-ATCTT 2030
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 236 gGlyCysLysLysLeu---TrpAlaValSerGlySerValPheLeuLeuGlyCysValAl 255
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2031 GGGTAAGCGCAAGACCGGCTTATCATTCGCGTGATTTATCATCGTTTGGGACTTATAGC 2090
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 255 a-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyG1 272
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2091 TTCATTTACAATCCGCTCTCAATAGG-----GGTATTGAATCTCCGGAGG 2135
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 272 y-----TyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMe 290
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

## RESULT 9

ABK74549

ID ABK74549 standard; DNA; 1584 BP.

XX

AC ABK74549;

XX

DT 13-AUG-2002 (first entry)

XX

DE Bacillus licheniformis genomic sequence tag (GST) #1840.

XX

KW Differential gene expression; genomic sequenced tag; GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.

OS Bacillus licheniformis.

PN WO200229113-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US31437.

XX 06-OCT-2000; 2000US-0680598.

XX 27-MAR-2001; 2001US-279526P.

XX (NOVO ) NOVOZYMES BIOTECH INC.

XX (NOVO ) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus  
 cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic  
 PT sequenced tag array

PS Claim 4; SEQ ID NO 1840; 200pp; English.

XX The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions.  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1584 BP; 389 A; 347 C; 395 G; 453 T; 0 other;

# Alignment Scores:

Pred. No.:	4,53e-64	Length:	1584
Score:	664.00	Matches:	173
Percent Similarity:	48.98%	Conservative:	114
Best Local Similarity:	29.52%	Mismatches:	177
Query Match:	23.90%	Indels:	122
DB:	24	Gaps:	16

US-09-868-987-14 (1-552) x ABK74549 (1-1584)

Qy	17	ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerLeuLysSerGly	36
Db	21	ATTCAAGGCACTTTACGATTTCAGAAAGCAAAACATCGGAGCATTTTAAATCGCGGA	80
Qy	37	AlaMetSerPheValProGluValLeuSerGluGlu-----ThrIleSerSerAsp	53
Db	81	GCTCTT-----CCTGTCAAACTCGTTGAAAAGATTTCGACATCCGTCGGAGCGCAG	131

Qy	54	LeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeu	73
Db	132	TTTGGGAGCAGCCTTTGAATCATACCGTGTTCGGGAATCATTTGTTATCTATATT	191
Qy	74	IleValLeuMetSerValTyrTyrArgPheGlyValIleAlaSerGlyAlaValLeu	93
Db	192	TTCTATTATGCTTCTTACTATCGCTTGCCTGCTGTTTATT-----CGGTCATT	242
Qy	94	LeuAsnLeuLeuIleTrpAlaLeuGln-----TyrLeuAspAlaProLeu	110
Db	243	ACGCTTTCGGTCTACATCTACATTACATGCGAGATATTCGACTGGATGCGGTTCTG	302
Qy	111	ThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsnVal	130
Db	303	ACCTTTCGGGATCGCGCTCTTATTTGGCGTCGAGATGGCGTTTCACCCACATC	362
Qy	131	LeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLysSerVal	150
Db	363	ATTACGTATGAACGCATAAAGAGAGCTGGAAGCTCGGCAATCGTCCGCTCGGCTTT	422
Qy	151	GluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThrValLeu	170
Db	423	AAAGCAGGAACAGAGATCGTTTGCACGATTTTGTATGCCAATATTCAGCAGTCTT	482
Qy	171	AlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu	190
Db	483	GCGGNAATCGTCTTTTCATATTTGGTACAAGCTCTGTCAAAGGCTTCGGACCATGCTG	542
Qy	191	IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMet	210
Db	543	ATCCTATCATTTTGACAAAGCTTTTATACAGCTGTCTCTCGCGCTTCCTGCTCGGC	602
Qy	211	Leu-----TrpMetAsnLysThrGln-----	217
Db	603	CTGCTGGTCCAAAGCGCTGGCTTGACCGGAAAAAGGCTGTGGGTGGCGCAAAAAA	662
Qy	218	-----HisThrGlnLeuHisMetMetAsnLysPheValGly	229
Db	663	GACATTTTGGATATCAGAAAGACAGACGAAATACGGAGCTCCGNAACCGTT-----	716
Qy	230	IleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPhe	249
Db	717	TCCAAATGGATTTCGTGGGAAGCGCAATGGTTCTTTGCATTTTCGGCGCTTCTCTG	776
Qy	250	LeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPhe	269
Db	777	GCGCTGGTCTTATCGTCTTGTGTAAAGTGAAC-----CTCGGCATCGATTTC	830
Qy	270	LysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGln	289
Db	831	TCAAGCGGA-----TCGACA	845
Qy	290	MetArgGlyLysValValHisLysLeu-----GlnGlu	300
Db	846	ATCGAAGTTCAAAGCGCATCATAGCTGACACCCAGCAGCTCGAAAAGACTTTGAACAA	905
Qy	301	AlaGlyLeu-----SerSerArgAspPheArgIleGln	311
Db	906	GTGCGGCTTGTATCTCCTGACTCGATCGCTTTTCAGCGCAAAAAAATCATCGCGTCGCC	965
Qy	312	ThrPhe-----GlySerSerGluLysIleLysIleTyrPheSerAsp	325
Db	966	AGTTTTCGCGCTGCCGATCAGAAAAGATCGCTGAAGTGAAGATTTTAAAGAC	1025
Qy	326	LysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSer	345
Db	1026	AAA-----	1028
Qy	346	TrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrClyAsnSerLys	365
Db	1029	-----TACGGAAGCGCAACCA	1043
Qy	366	ArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGln	385

```

Db 1044 AATGTCAGCACCGTATCGCTACCGTCGCGAAGAGCTGGCGAGAAATGCGCTTTATGCG 1103
Qy 386 AlathrileGlyLeuLeuGlyAlaLeuAlaLeuLeuLeuLeuValSerLeuArgPhe 405
Db 1104 GTGATCATCGCGTCAATCGGG-----ATCATTTTGTATGTCGATCGCGCTTT 1151
Qy 406 GluTrpGlnTyrAlaPheSerAlaValCysAlaLeuLeuHisAspLeuLeuAlaThrCys 425
Db 1152 GAATACAAGATGGCGATTGCGGCTATTACTTTCATTGCTGTATGAC----- 1196
Qy 426 AlaValLeuPheLeuAlaHisPheLeuLeuLysIleGlnIleAspLeuGlnAlaIle 445
Db 1197 GCGTATTTTCATTCATTCGCGTCTTCAGCCTGCAAGGCTTGAGTTGATGTACTTTCATT 1256
Qy 446 GlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuLeuIlePheAspArg 465
Db 1257 GCTCGGTCGTGAGATCATCGGCTATTTCGATAATGATACAATGTAAACGTTTGACAGG 1316
Qy 466 IleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuVal 482
Db 1317 ATCAGGAGCATATCAAAAGCGCAACCGAAACCTTCAGCGACTTGTCGATATTGTG 1376
Qy 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502
Db 1377 AATTGAGCTTGCAGCAGACGCTTCAGCGCTTCGATTAAATACGCTGCTGACTGCTGCTATT 1436
Qy 503 ValLeuLeuMetLeuLeuPheIleGlyCysSerSerValPheAsnPheAlaPheIleMet 522
Db 1437 GTCGTCAATTGCACATTCATCTTCGCGGCTGCATCGAATTCGAACTTCTCTGTGCGCTG 1496
Qy 523 ThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu 542
Db 1497 TTGTCGGACTTTTGAGCGGAGTATATCATCATCTACATCGCAGCTCAACTATGCTT 1556
Qy 543 PheMetValArgLysGlu 548
Db 1557 GTCTGGAAAGGCGCGAA 1574

RESULT 10
ABT15148
ID ABT15148 standard; DNA; 2292 BP.
XX AC ABT15148;
XX AC
XX DT 06-MAR-2003 (first entry)
XX DE Pathogen specific antigen related staphylococcal DNA SEQ ID No 534.
XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX KW autoimmune disease; HIV; hepatitis; gene; ds.
XX OS Staphylococcus sp.
XX PN WO200259148-A2.
XX PD 01-AUG-2002.
XX PF 21-JAN-2002; 2002WO-EP00546.
XX PR 26-JAN-2001; 2001AT-0000130.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Meinke A, Nagy E, Von Ahseu U, Klade C, Henics T, Zauner W;
XX PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
XX PI Tempelmaier B;
XX DR WPI; 2003-075410/07.
XX PT Identifying, isolating and producing hyperimmune serum-reactive

```

```

PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation -
XX Example 7; Page 245; 252pp; English.
XX The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens,
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This polynucleotide sequence represents
CC staphylococcal DNA relating to the method for identifying and producing
CC pathogen specific antigens of the invention.
XX SQ Sequence 2292 BP; 817 A; 332 C; 400 G; 739 T; 4 other;

```

```

Alignment Scores:
Pred. No.: 1-39e-62 Length: 2292
Score: 653.00 Matches: 166
Percent Similarity: 51.31% Conservative: 128
Best Local Similarity: 28.97% Mismatches: 199
Query Match: 23.51% Indels: 80
DB: 25 Caps: 13

```

US-09-868-987-14 (1-552) x ABT15148 (1-2292)

```

Qy 2 ValSerSerProLeuLeuAenValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21
Db 616 GTAGACCAACCTATTAAATCTAGTAGTGTGTAATTTTCAGGTGGCTTTCAGGAAAAA 675
Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe--- 40
Db 676 GGTGTTCGAAGACGCAACAAATAGCTGAGTTTAAATGCCGCTCATACCAGTTGAT 735
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
Db 736 TTAAGAAGAAATTTACTCTAACTCWNRDGTTGGTGCAACAATTTGGTCAAGATGCTCTT 795
Qy 61 GlnGlyIleSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValThr 80
Db 796 AAGACCATGTTGCTCAATTTAGGTATAGCATAATTAATTTATTTATTTATGCTGTTTC 855
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTrp 100
Db 856 TATCGTTTGGCTGCTGTTAGTTGCAATCATTCCTTAACCACTTATATTTAATTAACTTTA 915
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 916 GTCGCATTTCAATTTTCATATCAGGTGATTAACTCTACCTGGATGGCGGATAGTTTAA 975
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140
Db 976 GGTGTAGGTATGGCTGTCGATGCAATATCATATGATGAACGATATAAAGATGAACATA 1035
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 1036 AGAATTGGACGCGCTTAAACAAAGCGTATTTCAAAAGCAATAAAAGTTTCTTAACT 1095
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180

```

Db 1096 ATATTTGATTCACAACTTAACAACTGTCATCGTCAGCTGTCTTTCTTCTTTGGAGAA 1155  
 Qy 181 GlyProlLeuGlyPheAlaLeuThrLeuLeuLeuGlyPheSerSerMetPheThr 200  
 Db 1156 AGTTCAGTCAAGAGCTTCGCAACCATGTACTCTTAGGTATTTTAATGATATTTGTAACC 1215  
 Qy 201 AlaLeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1216 GCAGTATTCATCAAGAGGGTGTATCATCTACTGTGTATCTTCAAACTTCTTTAAAAAA 1275  
 Qy 212 -----TrpMetAsnLysThrGlnHisThrGlnLeuHisMetAsnLysPheValGly 229  
 Db 1276 CAATACCTGGTTATTGGTTTAAAGAGAGATAGACATGATATTAATGAAGGTAAAGAT 1335  
 Qy 230 IleLysHisAsp-----PheLeuArgGlyCysLysLys 240  
 Db 1336 GTA---CATGATTTTAAAAACATCATATCAAGAGTTAACTTTGTTAAATTAGCTAAGCCA 1392  
 Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 1393 CTTATTCACCTAGTATTTAAATTTGTAATTTGTTGTTGTTGTTATCATTTCAATATTAAAA 1452  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280  
 Db 1453 TTAAAC-----TTAGGTATTGATTTCTCATCCGGAACAAGAGCAGATATTCATCTAAA 1506  
 Qy 281 GluHisGlyLysSerAspValAlaGlnMetArgGlyLysValHisLysLeuGlnGlu 300  
 Db 1507 AAT-----GCTATAACAACAAGCAGAGTTGAGAAAACTGTAAAAATCA 1548  
 Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
 Db 1549 GTTGATTTGACACAGAT-----CAATACAGATTAATGTGTGGAATAAA----- 1596  
 Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLys 340  
 Db 1597 -----AATGCCACAGTTCAGITTTAAAAAAGATTTATCACGT 1632  
 Qy 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360  
 Db 1632 ----- 1632  
 Qy 361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1633 GAGGAACACATAAATAAGTGTCTAAGTGCAATCTGAAATTTGGAGATAATCCACAAATT 1692  
 Qy 374 -----ValSerSerLysLeuSerLysMetArgTyrGlnAlaThrIleGlyLeuLeu 391  
 Db 1693 AATACCGTTTCACCTCTCATAGGCCAAGAGCTAGCTAAATAATGCTGTAACGTAAATA 1752  
 Qy 392 GlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPhe 411  
 Db 1753 CTTGCTTCTTAGGCATATTATCTATGTTTCACTAAGATTTGAATGCGTATGGGTCTA 1812  
 Qy 412 SerAlaValCysAlaLeuIleHisAspLeuAlaThrCysAlaValLeuPheIleAla 431  
 Db 1813 TCTTCTGACTGCTGATTTATACATGACGTATT-----ATCACTATTGCA 1857  
 Qy 432 HisPheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrVal 451  
 Db 1858 ATCTTTAGTTGTTTACAGTAGAGTAGATTAACTTTATTGTCAGCAGATTAACTATC 1917  
 Qy 452 LeuGlyTyrSerLeuAsnAsnThrLeuIlePheAspArgIleArgGluAsp---Arg 470  
 Db 1918 GTTGCTTATTCAATCAATGATACAAATCGTAACCTTTCACCGGTGTCGAGAAAAATCTGCAT 1977  
 Qy 471 GlnAlaAsnLeuPheThr-----ProMetHisValLeuValAsnAspAlaLeuGln 487  
 Db 1978 AAGTTAAAGTAATTACGCATCTACTGATCAAAATGATGATAGTCAACCGCTCTATTAGA 2037  
 Qy 488 LysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeu 507  
 Db 2038 CAAATATGACACGTTCTTATTAAATACAGTGTGTTGACTGTAGTTGTAGTTGTGCAATA 2097

Qy 508 LeuPheIleGlyClySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeu 527  
 Db 2098 TTAATATTAGGTGACCAACAATATTAATTTCTCTTTAGCATTTACTAATTTGATTATTA 2157  
 Qy 528 LeuGlyThrLeuSerSerLeuTyrIleAlaProProLeu 540  
 Db 2158 TCTGGTGTTATCTCGTCAATTTTCATTTGCTGTACCATTA 2196

## RESULT 11

AAH52892  
ID AAH52892 standard; DNA; 2226 BP.

XX AAH52892;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1177.

XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
KW vaccination; endocarditis; ds.

XX Staphylococcus epidermidis.

PN WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX ) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX P-PSDB; AAG82042.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
XX useful for vaccinating against infections, e.g. endocarditis -

XX Claim 8; Page 337-338; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH53091 to  
 CC AAH5098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 2226 BP; 796 A; 320 C; 387 G; 723 T; 0 other;

## Alignment Scores:

Pred. No.: 7,116-62 Length: 2226  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79

[illegible]

PI Doucette-Stamm LA, Bush D;  
 XX WPI; 2002-381255/41.  
 DR P-PSDB; ABP38844.  
 XX

Novel isolated nucleic acid encoding a *Staphylococcus epidermidis* polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 PS Disclosure; SEQ ID 852; 267pp; English.

XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the CC USPTO web site.

XX  
 SQ Sequence 2301 BP; 819 A; 333 C; 403 G; 746 T; 0 other;

Alignment Scores:  
 Pred. No.: 7,49e-62 Length: 2301  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79  
 DB: 24 Gaps: 12

US-09-868-987-14 (1-552) x ABN91389 (1-2301)

Qy 2 ValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21  
 Db 625 GTAGACCAACCTATTAAATCTAGTAGTCTGCTGAAATTTTCAGGTGGCTTCAATGGGAAAAA 684

Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 685 GGTGTGAAGAACGAAACAAATAGCTGAGTTATTAATAGCCGGCTCATTACCAAGTTGAT 744

Qy 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61  
 Db 745 TTAAAGAAATTTACTCTAACTCTGTTGGTGCAACAATTTGCTCAAGATGCTCTGTGAAG 804

Qy 62 GlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 81  
 Db 805 ACCATGTTGCATCAATTTGATGATAGCTAGCTAATTAATTTATTTATGCTTGGTTCTAT 864

Qy 82 ArgPheGlyGlyValIleAlaSerGlyValValLeuLeuLeuAsnLeuLeuIleTrpAla 101  
 Db 865 CGTTTGGCTGTTTGTAGTTGCAATCATGTCCTTAAACCACTTATATTTTAACTTTAGTC 924

Qy 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121  
 Db 925 GCATTCAATTTTCATACAGGTGATTAACTCTACCTCGATGGATGGCGCATTTAGGT 984

Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 985 GTAGGTATGGCTGTCGATGCAATATCATATGATGATGAAAGTATTAAGATGAAGA 1044

Qy 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161  
 Db 1045 ATTGGACGACGCTTAAACAGCGGATTCAAAAGCAAAATAAAGTTTCATTCTTAACTATA 1104

Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181  
 Db 1105 TTGATTCCAACTTAAACAACCTGTCATCGCTGCAGCTGCTTTCTCTTTGGAGAAAGT 1164

Qy 182 ProLeuLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201  
 Db 1165 TCAGTCAAGGCTTCGCAACCACTGTTACTCTTAGGTATTTTAATGATATTTGTAACCGCA 1224

Qy 202 LeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1225 GTATTCCTTCAAGAGGGTTGTTATCATTTACTACTGGTATCTTCAAACTCTTTTAAAAACAA 1284

Qy 212 ---TrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230  
 Db 1285 TACTGGTTATTTGGTGTAAAGAGATGGATAGACATGATATTAAATGAAGTAAAGATGTA 1344

Qy 231 LysHisAsp-----PheLeuArgGlyCysLysLysLeu 241  
 Db 1345 ---CATGATTTAAAAACATCATATGAAGAGTTAAACTTGTAAATTAGTAGTAAGCCACTT 1401

Qy 242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261  
 Db 1402 ATTTCTACTTAGTATTTTAAATTTGTTGTTGATATCATTTCAATATTTAAATTA 1461

Qy 262 AsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGlu 281  
 Db 1462 AAC-----TTAGTATTGATTTCTCATCGGAAACAGACAGATATTCAATCTAAAAAT 1515

Qy 282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301  
 Db 1516 -----GCTATAACACAGACAGGTTGAGAAAACTGTAAAAATCAGTT 1557

Qy 302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle 321  
 Db 1558 GGATTTGGAAACACAGAT-----CAATACAGATTAATGGTAGTGGAAATAA----- 1602

Qy 322 TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeu 341  
 Db 1603 -----AATGCCACAGTTCACTTTAAAAAAGATTTATCACGT--- 1638

Qy 342 ThrIleMetSerTrpArgTyrCysGlyIleValValAlaGAsnArgProArgPheLeuTyr 361  
 Db 1639 -----GAG 1641

Qy 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1642 GAAGACAAATAAATAAGTGAAGTGAATTTGAGATTAATCCACAAATAAT 1701

Qy 374 ---ValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1702 ACCGTTTCACTCTCATAGCCCAAGAGTAGCTAAAAATGCTCTAACTGCATTAACTT 1761

Qy 393 AlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSer 412  
 Db 1762 GCTTCTATAGGCATTATTATCTATGTTTCAATGAATTTGAATGGGTATGGCTCTATCT 1821

Qy 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1822 TCTGTACTTGCATTATTATCATGACGATTTT-----ATCATCATTTGCAATC 1866

Qy 433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 1867 TTTAGTTTGTGTAGTAGTAGTAGATTAAACATTTATTTCAGCAGTATTAACTATCGTT 1926

Qy 453 GlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp---ArgGln 471  
 Db 1927 GGTATTATCAATCAATGATACAAATCGTAACCTTTCGACCCTGTTCGAGAAAAATCTGCATAA 1986

Qy 472 AlaAsnLeuPheThr-----PrometHisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 1987 GTTAAAGTAATTACGCATATCTGATCAAAATTTGATGATATAGTCAACCGCTCTCTATTAGCAA 2046

Qy 489 ThrPheSerArgThrValIleMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 2047 ACTATGACACAGTTCTATTAAATACAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2106

Qy 509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 Db 2107 ATATTAGGTGCGCAACCAATATTTAAATTTCTCTTAGCATTTACTAATTTGGATTATTATCT 2166

QY 529 GlyThrLeuSerSerLeuTyrIleAlaProProLeu 540  
 Db ||||| : : : : : |||||  
 2167 GGTGTAATTCGTCAATTTCATTGCTGTACCATTA 2202

RESULT 13  
 AAH54097/C  
 ID AAH54097 standard; DNA; 3511 BP.  
 AC AAH54097;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3461.  
 XX  
 KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 KW vaccination; endocarditis; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX  
 PA (GLAX) GLAXO GROUP LTD.  
 XX  
 PI Kimerly WJ;  
 XX  
 DR WPI; 2001-316495/33.  
 XX  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 8; Page 1027-1028; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC -The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 3511 BP; 1155 A; 610 C; 483 G; 1263 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,45e-61 Length: 3511  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.23% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79  
 DB: 22 Gaps: 12

US-09-868-987-14 (1-552) x AAH54097 (1-3511)

QY 2 ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21  
 ||||| : : : : : |||||  
 1920 -----CAG 1918

2934 GTAGACCAACCTATTAATTTCTAGTAGTGTGTTGAAATTTTCAGGTGGCTTCAATGGGAAAAA 2875  
 QY 22 ThrHisargGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db ||||| : : : : : |||||  
 2874 GGTGTTGAAGAACGAAACAAATAGCTGAGTTATTAATAGCCGCTCATTTACCAAGTTGAT 2815  
 QY 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61  
 Db ||||| : : : : : |||||  
 2814 TTAAAGAGAAATTTACTCTAACTCTGTTGGTGCACAAATTTGGTCAAGATGCTCTTGATAAG 2755  
 QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 81  
 Db ||||| : : : : : |||||  
 2754 ACCATGTTTGCATCAATTTAGGTATAGCATTAATTTATTTATTTATTTGTTGTTCTAT 2695  
 QY 82 ArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuTrpAla 101  
 Db ||||| : : : : : |||||  
 2694 CGTTTGCTGTTTGTGTTGCAATCATTCCTTAACCACTTATATTTATTTAACTTTAGTC 2635  
 QY 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121  
 Db ||||| : : : : : |||||  
 2634 GCATTCAATTTTCATATCAGGTGTATTAACCTCTACCTGGATGGCGCATTTAGGT 2575  
 QY 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db ||||| : : : : : |||||  
 2574 GTAGGTATGGCTGCTGATGCCAATATCATAATGTATGAACGTATTAAGATGAACATAAGA 2515  
 QY 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161  
 Db ||||| : : : : : |||||  
 2514 ATTGGACGCGCTTAAACCAAGCGTATTCAAAAGCAAAATAAAGTTCAATCTTAACATATA 2455  
 QY 162 PheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuLeuAspThrGly 181  
 Db ||||| : : : : : |||||  
 2454 TTTCATTTCCAACTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2395  
 QY 182 ProIleLysGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPheThrAla 201  
 Db ||||| : : : : : |||||  
 2394 TCAGTCAAAGGCTTCGCAACCATGTTACTCTCTAGGTATTTTAAATGATATTTGTAACCGCA 2335  
 QY 202 LeuPheMetThrLysPhePheMetLeu----- 211  
 Db ||||| : : : : : |||||  
 2334 GTATTTCTTACAGAGGGTGTGTTATCATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2275  
 QY 212 ---TrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230  
 Db ||||| : : : : : |||||  
 2274 TACTGGTATTGCTGTTGTTGAAGAGATAGATGATATTAATGAAGGTAAGATGCTA 2215  
 QY 231 LysHisAsp-----PheLeuArgGlyCysLysLeu 241  
 Db ||||| : : : : : |||||  
 2214 ---CATGATTTAAACACATCATATGAAGGTTAAACCTTTGTTAAATTTAGTAAAGCCACTT 2158  
 QY 242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261  
 Db ||||| : : : : : |||||  
 2157 ATTTCACTTAGTATTTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2098  
 QY 262 AsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGlu 281  
 Db ||||| : : : : : |||||  
 2097 AAC-----TTAGTATTGATTTCTCATCCGGAACAGACAGATATTTCAATCTTAAAT 2044  
 QY 282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301  
 Db ||||| : : : : : |||||  
 2043 -----GCTATAACACAGACAGGTTGAGAAACCTGTAATAATCAGTT 2002  
 QY 302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLysValle 321  
 Db ||||| : : : : : |||||  
 2001 GGATTGGAACACAGAT-----CAAATACAGATTAAATGTTAGTGAATAA----- 1957  
 QY 322 TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeu 341  
 Db ||||| : : : : : |||||  
 1956 -----AATGCCACAGCTTCAGTTTAAAGAAAGATTTATCAGT---- 1921  
 QY 342 ThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyr 361  
 Db ||||| : : : : : |||||



QY 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1917 GAAGACAATAAATTAAGCTGAAGTGAATCTGAATTTGGAGATAATCCACAATAAT 1858  
 QY 374 ---ValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1857 ACCGTTTCACCTCATAGGCCAAGAGCTAGTAAATGCTGTAACGTCAATTAATCTT 1798  
 QY 393 AlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSer 412  
 Db 1797 GCTTCTATAGCAATTATTATCTATGTTTCACTAAGATTTGAATGGGTATGGTCTATCT 1738  
 QY 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1737 TCTGTACTTGCAATTATTACATGACGTATTT-----ATCATCATTTGCCAATC 1693  
 QY 433 PhePheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 1692 TTTAGTTTGTAGATTAGATGATTTTACATTTATTGACGAGTATTAACCTATCGTT 1633  
 QY 453 GlyTyrSerLeuAsnThrLeuIlePheAspArgIleArgGluAsp---ArgGln 471  
 Db 1632 GGTATTCAATCAATGATACAAATCGTAACCTTTCGACCGTGTTCGAGAAATCTGCATAA 1573  
 QY 472 AlaAsnLeuPheThr-----PromethisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 1572 GTTAAGTAATTTACGCATCTACTGATCAAAATTCATGATATAGTCAACCGCTCTATTAGACAA 1513  
 QY 489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 1512 ACTATGACACGTTCTATTATTAACAGTGTGACTGTGATGTTGATGTTGATGTTGATGTT 1453  
 QY 509 PheIleGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 Db 1452 ATATTAGTGCCACCAACAATTTAATTTCTTTAGCATTACTAATTTGGATTATTATCT 1393  
 QY 529 GlyThrLeuSerSerLeuTyrIleAlaProLeu 540  
 Db 1392 GGTGATTCTCGTCAATTTTTCATTGCTGTACCATT 1357  
 RESULT 14  
 AAH54228/c  
 ID AAH54228 standard; DNA; 3705 BP.  
 XX AAH54228;  
 AC  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3592.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX  
 FA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Kimmerly WJ;  
 XX  
 DR WPI; 2001-316495/33.  
 XX  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX useful for vaccinating against infections, e.g. endocarditis -

PS Claim 8; Page 1183-1184; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX  
 SQ Sequence 3705 BP; 1234 A; 621 C; 509 G; 1341 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,57e-61 Length: 3705  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79  
 DB: 22 Gaps: 12

US-09-868-987-14 (1-552) x AAH54228 (1-3705)

QY 2 ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21  
 Db 2211 GTAGACCAACCTATTAACTTAGTAGTGTGTAATTCAGGTGGCTTCAATGGGAAAAA 2152  
 QY 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 2151 GGTGTGAAGAGCGAAACAAATAGCTGAGTTATTAAATGCCGCTCAATACCACTTGAT 2092  
 QY 42 ProGluValLeuSerGluThrIleSerSerAspLeuLysGlyLysGlnCysThrGln 61  
 Db 2091 TTTAAAGAAATTTACTCTACTGTTGTGTGCACAAATTTGGTCAAGATGCTCTTTGATAAG 2032  
 QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTyr 81  
 Db 2031 ACCATGTTGTCATCAATTTAGTAGTAGCATTAATTTATTATTATTTGCTGTTCTAT 1972  
 QY 82 ArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAla 101  
 Db 1971 CGTTTGCCTGGTTAGTTCGAATCATTTGCTTAAACCACTATATATTATTAACTTTAGTC 1912  
 QY 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaClyIleValLeuAla 121  
 Db 1911 GCATTCAATTTTCATATCAGGTGATTAACTTACCTGCGATTTGGCGCATTTAGTTTAGGT 1852  
 QY 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 1851 GTAGGTATGGCTGTCGATGCCAATATCATATGATGATGATTAAGATTAAGTAAAGTAA 1792  
 QY 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161  
 Db 1791 ATTGACGACGCTTAAACAAAGCTATTCAAAGCAATATAAAAGTTTCATTTCTTAACATA 1732  
 QY 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuPhePheLeuAspThrGly 181  
 Db 1731 TTTGATTCCAACTTAAACACTGTCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672  
 QY 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201



Db 1671 TCAGTCAAGGGCTTCGCAACCATGTTACTCTTAGGTATTTAAATGATATTTGTAACCGCA 1612  
 Qy  
 Db 202 LeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1611 GATATTCATTCAAGAGGGTGTATTCATTTACTGGTATCTTCAAACTTCTTTAAAAACA 1552  
 Qy 212 ---TTPMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230  
 Db 1551 TACTGGTATTGGTGTAAAGAGGAGGATAGACATGATATTAAGAGGTAAGATGTA 1492  
 Qy 231 LysHisAsp-----PheLeuArgGlyCysLysLysLeu 241  
 Db 1491 ---CATGATTTAAAAACATCATATGAAAGGTTAAACTTTTGTAAATTAGCTAAGCCACTT 1435  
 Qy 242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261  
 Db 1434 ATTTCACATTAGTATTTAATGTAATTAATGTTGTTGATTATCATTTCAATATTTAAATTA 1375  
 Qy 262 AsnSerValLeuGlyMetAspPheLysGlyGlyTyrrAlaPheThrPheAsnProLysGlu 281  
 Db 1374 AAC-----TAGGTATTGATTTCTCATCCGGAACAAGCAGATATTCATCTTAAAAAT 1321  
 Qy 282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301  
 Db 1320 -----GCTATAACAACAAGCAGGTTGAGAAAACTGTAATAATCAGTT 1279  
 Qy 302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle 321  
 Db 1278 GGATTTGGAACCCAGAT-----CAATACAGATTAATGTTAGTGGAAATAA----- 1234  
 Qy 322 TyrPheSerAspLysAlaLeuSerTyrrLysGlnIleArgAlaSerLeuLysLeu 341  
 Db 1233 -----AATGCCACAGTTTCAGTTTAAAAAGATTTATCAGT--- 1198  
 Qy 342 ThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyr 361  
 Db 1197 -----GAG 1195  
 Qy 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1194 GAAGACAATAAATTAAAGTGCTAAGTGAAATCTGAATTTGGAGATAAATCCACAATAAT 1135  
 Qy 374 ---ValSerSerLysLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1134 ACCGTTTCACCTCTCATGGCAAGAGCTAGCTAAATAATGCTGAATGCAATTAATCTT 1075  
 Qy 393 AlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSer 412  
 Db 1074 GCTTCTATAGGCATTATTATCTATGTTTCACTAAGATTTGAATGCGGTATGCTATCT 1015  
 Qy 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1014 TCTGTACTTGCATTATTATCATGACGTATTT-----ATCATCATTCGAATC 970  
 Qy 433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 969 TTTAGTTGTTTGTAGATTAGAGTAGATTAACTATTATTCGCGCAGTATTAACTATCTGT 910  
 Qy 453 GlyTyrSerLeuAsnAsnThrLeuIlellePheAspArgIleArgGluAsp---ArgGln 471  
 Db 909 GGTATTCAATCAATGATACATCTGAATTTTCGACCGGTTCGAGAAAAATCGCATAA 850  
 Qy 472 AlaAsnLeuPheThr-----ProMetHisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 849 GTTAAGTAATTAACCATCTAGTCAAAATTTGATGATATAGTCAACCGCTCTATTAGACA 790  
 Qy 489 ThrPheSerArgTyrValMetThrThrAlaThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 789 ACTATGACACGGTCTTATTAAATACAGTGTGCTGTAGTTGTAGTTGTCGAATATA 730  
 Qy 509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyLeuLeu 528  
 Db 729 ATATTAGGTGCAACCAATAATTTAATTTCTTCTTAGCATTACTAATTTGGATTATATCT 670

Qy 529 GlyThrLeuSerSerLeuTyrrIleAlaProProLeu 540  
 Db 669 GGTGATTCTCGTCATTTTTCATTTGCTGTACCAATTA 634

## RESULT 15

AAA81457/C  
 ID AAA81457 standard; DNA; 49646 BP.  
 XX  
 AC AAA81457;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis partial DNA sequence gnm\_5 SEQ ID NO:5.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US235573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlati V;  
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlati V;  
 PI Rappuoli R, Pizza M;  
 XX  
 DR WPI; 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 PS Claim 7; Page 274-288; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SQ Sequence 49646 BP; 12331 A; 14486 C; 11862 G; 10966 T; 1 other;  
 Alignment Scores:

```
Pred. No.: 1.25e-56 Length: 49646
Score: 618.50 Matches: 160
Percent Similarity: 51.19% Conservative: 140
Best Local Similarity: 27.30% Mismatches: 209
Query Match: 22.26% Indels: 77
DB: 21 Gaps: 15

US-09-868-987-14 (1-552) x AAA81457 (1-49646)

Qy 1 MetValSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGly 19
Db 31089 GTTGTAAACCGCGCGGTTATCCGTAATCCGCGGACGCGTGGAATTTCCGGA 31030
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 31029 AGCATGACGACGCGGAACCAATGATACGCTTTGCTGTTGCGTTCCTTCGCC 30970
Qy 40 PheValProGluValLeuSerGluGluThrLysSerAspLeuGlyLysLysGlnCys 59
Db 30969 GCACCGATGACAGATTGTCGAAGAACCTACCATCGCTCGCTCTTTGGGTAAAGGAACATC 30910
Qy 60 ThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerVal 79
Db 30909 GAAAAAGGCTTCCATTCGACTTTATGGGGTTTCCCATCGTTGCTGCAATTCATCGTGGT 30850
Qy 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIle 99
Db 30849 TACTATCGTCGATGGGTTCTTTTCTACCATTCGATTGAGTGCCAACTACTGTTCTTA 30790
Qy 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119
Db 30789 ATCGGTATTTGTCGCCATCAGCAACGTTGACGTTACCGGGTATGCGCGCTGGCG 30730
Qy 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGlu 139
Db 30729 TTGACTTTGGGTATGGCAATCGACTCCAACTCTTGATTAAACGAACGCTATCCGCGAAGA 30670
Qy 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159
Db 30669 TTGCGTCCGCGGTCGCCCGCAGCAGCATCAATCTCGGTTTCCAAACGATGGCG 30610
Qy 160 AlaIlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAsp 179
Db 30609 ACCATTGTCGATTCCAACTCGACTTCGCTGATTCGCGCTATCGCGCTTTGGTATTCGGT 30550
Qy 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPhe 199
Db 30549 TCCGCGCGGTACCGGGTTTTCGGGTCGTACACTGTTTGGGTATTCGACTTCGATGTAT 30490
Qy 200 ThrAlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThr 219
Db 30489 TCATCCGTCGTCGATTCCGCGGCTTGGTCAATCTGTGGTACGACGACGACGCAAAATG 30430
Qy 220 GlnLeuHisMetMetAsnLysPheValGly----- 229
Db 30429 CAG-----AATATTTCATTGTTGCGGTGGAAGCCGAAAGCCGAAATGGCA 30382
Qy 230 ----- 236
Db 30381 GGAGGCAAGGAGTAAGCTATGGAACCTTTTAAATCAACCGCATATTCGTTTATGAGC 30322
Qy 237 GlyCysLysLysLeuTrpAlaValSer-----GlySerValPheLeuLeu 251
Db 30321 TACGCGAACTGACGACCTTCATTTCGTTGATTACGTTTATCGTGGCGGTCTTTTG 30262
Qy 252 GlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGly 271
Db 30261 GTTACCAGAGGTCGAAATTC-----TCTGTGCAATTTACCGGC 30223
Qy 272 GlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMetArg 291
Db 30222 GGT-----ACGGTAATGGAAGTCCAATATATCAGCAG 30193
```

```
Qy 292 GlyLysValValHisLysLeuGlnGluAla-----GlyLeuSerSerArgAspPheArg 309
Db 30192 GGTGCGGATGTCATTAAGATGCGCGAAGCGCTCGATACGCTGAAATAGGTGATGATACAG 30133
Qy 310 IleGlnThrPheGlySerSerGluLysIleLysIleLysIleLysIleLysIleLysIle 329
Db 30132 GTTCAGGCATTGGGTACGAACAACACATCATCGCTCGCTCGCAACAAA----- 30082
Qy 330 TyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyrCys 349
Db 30081 -----GAAGGTGTTACTTCCGCACAGTTGTCCAATCAGGTTATGGAT----- 30040
Qy 350 GlyIleValIleValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLys 369
Db 30039 ---TTGCTGAAAAAAGACAGTCCGACGTT-----ACCTTGCGCAAGTCGAA 29995
Qy 370 PheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGly 389
Db 29994 TTT-----ATCGCGCCGCAAGTCGGTAGGAATTTGTAAGTAATGATGATGGCT 29944
Qy 390 LeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTyrGlnTyr 409
Db 29943 TTAGGTTTGTCTGTTATCGGCATCATTTTACCTGTCGATGCTTTTGAATGCGGTTT 29884
Qy 410 AlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPhe 429
Db 29883 GCGTATCTGCCATTATCGCCAATATGCAGCATC-----GTGATTATT 29839
Qy 430 IleAlaHisPhePheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMet 449
Db 29838 CTCGGTCTGTTTCCCTTCTTCCAAATGGAAATTTTCGCTGACCGCTTGGCGGCTATCCTT 29779
Qy 450 ThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp 469
Db 29778 GCGTATTGGGCTATTCTGTGAACGAATCCGTCGCTGCTTCGACCGTATCCGTAAC 29719
Qy 470 -----ArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGln 487
Db 29718 TTCCGCAAGCCGCGATGCGCGACATGCGTCCGCGAAGTCATCCACACGCGATTACC 29659
Qy 488 LysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeu 507
Db 29658 GCAACGATGAGCGCGACCATCATTCACCGGTTTCGACGAGCGGATGCTGCTATCCATG 29599
Qy 508 LeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeu 527
Db 29598 CTGGTGTTCGGCGGTGCGGCTTTTCGACGCGCTTTTCTATGGCGTTGACCATTTGCGATCGTG 29539
Qy 528 LeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuLeuLeuLeuLeuLeu 546
Db 29538 TTCGGCATTTATTCTTCGCTATTGTTGTCGACCGCGCTTCTGCTAATGTTGCGTTTGAGC 29479
Qy 547 LysGluAsnArgSerLys 552
Db 29478 CGCGACAATATCGGTAAA 29461
```

Search completed: October 28, 2003, 16:14:11

Job time : 2871 secs